Locally Finite Languages

Olivier Finkel

LOCALLY FINITE LANGUAGES

Olivier Finkel
Equipe de Logique Mathématique
CNRS URA 753 et Université Paris 7
U.F.R. deMathématiques, 2 Place Jussieu 75251 Paris cedex 05, France.
E Mail: finkel@logique.jussieu.fr

Abstract

We investigate properties of **locally finite languages** introduced by J-P. Ressayre in [Res88]. These languages are defined by **locally finite sentences** and generalize languages recognized by automata or defined by monadic second order sentences. We give many examples, showing that numerous **context free** languages are locally finite. Then we study **closure properties** of the family *LOC* of locally finite languages, and show that most undecidability results that hold for context free languages may be extended to locally finite languages. In a second part, we consider an extension of these languages to **infinite and transfinite length** words. We prove that each α-language which is recognized by a **Büchi automaton** (where α is an ordinal and $\omega \leq \alpha < \omega^\omega$) is defined by a locally finite sentence. This result, combined with a preceding one of [FR96], provides a generalization of Büchi's result about decidability of monadic second order theory of the structure $(\alpha, <)$.

1 Introduction

In the sixties J.R. Büchi established some ties between monadic second order logic and finite automata, [Büc60]. When a word over a finite alphabet is considered as a structure in a natural manner, a finitary language is recognized by a finite automaton if and only if it is the class of finite models of a monadic second order sentence. And the distinction between first order and second order sentences is here very significant: first order sentences define the important class of star free languages, [Pin96] [Par83].

In 1974 R. Fagin proved that the class **NP** is characterized by existential second order formulas, [Fag93]. Since then, many complexity classes like **P, PSPACE, LogSPACE, NLogSPACE,** have been proved to be characterized by different versions of second order logic with the use of, for example, fixed point operators or transitive closure operators [Tho96]. This led to the now called descriptive complexity theory, (see [Tho96], [Fag93], [Imm87], [EF95] for more results and references). A logical characterization of context free languages has been recently found by C. Lauteman, T. Schwentick and D. Therien, involving second order quantifications over some special binary relations called matchings, [LST94]. A logical characterization for Petri Net languages had been found by M. Parigot and E. Pelz, [PP85].

To prove the decidability of the monadic second order theory of one successor over the integers, and further over the structure $(\alpha, <)$, for α a countable ordinal, J.R. Büchi and D. Siefkes studied ω-languages (and α-languages) which are recognized by finite automata reading infinite (and transfinite) length words. An α-language is shown to be accepted by a finite automaton iff it is defined by a monadic second order formula, [Büc62] ,[BS73]. Since then the expressive power of other logics, such as temporal logics, over infinite words has been studied, and many results have been stated. See [Tho90], [Tho96], [Pin96], [PP98], [Sta97], [LT94] for more results and references.

The case of transfinite length word languages appear in the theory of traces for the modelization of concurrency [DR95] and in the work about timed automata when one consider that an infinite number of actions may happen during a finite period of time [AD94], [BP97]. The languages of transfinite length words which are recognized by finite automata have been recently studied by N. Bedon in [Bed96], [Bed98] where the extension of the equivalence between first order sentences and star free languages is shown.

To extend these results, J-P. Ressayre, in order to apply ideas and machinery of model theory to the study of formal languages, introduced in [Res88] the locally finite sentences and the locally finite and free locally finite languages. He showed that these latter languages may be seen as a class of finite structures satisfying a strong finitary analogue to the properties of a universal and complete class of first order structures equipped with elementary embeddings.

Locally finite sentences are first order, but they define locally finite languages, via existential quantifications over relations and functions which appear in the locally finite sentence. And these second order quantifications are much more general than the monadic ones as the following results show:

- Each rational language is locally finite, [Res88].
- Each quasirational language is locally finite, Theorem 2.21 below.
- Each ω-language which is accepted by a Büchi automaton is a locally finite ω-language, [Fin89] and Theorem 5.7 below.
- And there exist much more locally finite (ω-)languages, see section 2 below and [Fin89],[Fin99].

But syntaxic and semantic restrictions (a locally finite sentence is equivalent to a universal one and its models satisfy simple structural properties) make locally finite ω-languages keep important properties of rational ω-languages. The pumping lemma of [Res88] is an analogue for each locally finite ω-language of the property : " A rational ω-language is non empty if and only if it contains an ultimately periodic word ". And this lemma implies in a similar manner the decidability of the emptiness problem for locally finite ω-languages. And similar results exist for α-languages, where α is an ordinal $< \omega^\omega$, [FR96].
And other pumping lemmas based on the existence of indiscernables in a model imply decidability of problems like: is a finitary locally finite language infinite?

We study here first the finitary locally finite languages.
In section 2, we give many examples. The question to know whether every context free language is locally finite remains open contrary to that is asserted in [Res88]. But Dyck languages and quasirational languages as many other context free languages are locally finite. We then study two hierarchies which are located between context free and context sensitive languages: there are locally finite languages at each level of these hierarchies.

In section 3, classical closure properties are investigated. Locally finite languages are closed under union, concatenation product and star operation [Res88]. We show that the class *LOC* of locally finite languages is closed under substitution, morphism, alphabetic inverse morphism, but not under intersection, complementation, intersection with a rational language, inverse morphism. These latest results are obtained by the use of the notion of rational cone.

In section 4, numerous undecidable problems for context free languages are shown to remain undecidable when locally finite languages are considered. In particular, given an alphabet Σ containing more than two letters, it is undecidable to determine, for arbitrary given locally finite languages $L(\varphi)$ and $L(\psi)$ over Σ, whether $L(\varphi) \cap L(\psi)$ is empty, infinite, rational, context free or locally finite, whether $L(\varphi) \subseteq L(\psi)$, whether

$L(\varphi) = L(\psi)$, whether $L(\varphi)$ is rational, whether ${}^cL(\varphi)$, the complement of $L(\varphi)$, is empty, infinite, rational, or locally finite.

In section 5 and 6, we consider an extension of locally finite languages to infinite and transfinite length words. We show that every α-language which is recognized by a Büchi finite automata, (with $\omega \leq \alpha < \omega^\omega$), is a locally finite α-language. This gives a new decision algorithm for the emptiness problem for Büchi α-languages. And this permits us to consider other new decidability results based on the use of indiscernables in a model. This shows also that the expressive power of the formulas $\exists R_1...R_k\varphi(R_1...R_k)$, where $\varphi(R_1...R_k)$ is locally finite in the signature $\{<, R_1, ..., R_k\}$, is stronger than that of the monadic second order formulas while keeping decidability properties for the structure $(\alpha, <)$, $\omega \leq \alpha < \omega^\omega$. Hence this provides a generalization of Büchi's result.

2 Examples of locally finite languages

2.1 First definitions

We briefly indicates now some basic facts about first order logic and model theory. More information may be found in textbooks, like [CK73] or [Sch67].

We consider here formulas of first order logic. The language of first order logic contains (first order) variables x, y, z, ... ranging over elements of a structure, logical symbols: the connectives $\wedge$ (and), $\vee$ (or), $\rightarrow$ (implication), $\neg$ (negation), and the quantifiers $\forall$ (for all), and $\exists$ (there exists), and also the binary predicate symbol of identity $=$.
A signature is a set of constant, relation (different from $=$) and function symbols. we will consider here only finite signatures.

Let Sig a finite signature. We define first the set of terms in the signature Sig which is built inductively as follows:

1. A variable is a term.
2. A constant symbol is a term.
3. If F is a m-ary function symbol and $t_1, t_2, ..., t_m$ are terms, then $F(t_1, ..., t_m)$ is a term.

We then define the set of atomic formulas which are in the form given below:

1. If t_1 and t_2 are terms, then $t_1 = t_2$ is an atomic formula.
2. If $t_1, t_2, ..., t_m$ are terms and R is a m-ary relation symbol, then $R(t_1, ..., t_m)$ is an atomic formula.

Finally the set of formulas is built inductively from atomic formulas as follows:

1. An atomic formula is a formula.
2. If φ and ψ are formulas, then $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \rightarrow \psi$ and $\neg\varphi$ are formulas.
3. If x is a variable and φ is a formula, then $\forall x\varphi$ and $\exists x\varphi$ are formulas.

An open formula is a formula with no quantifier.
We assume the reader to know the notion of free and bound occurrences of a variable in a formula. Then a sentence is a formula with no free variable.
A sentence in prenex normal form is in the form $\varphi = Q_1x_1...Q_nx_n\varphi_0(x_1, ..., x_n)$, where each Q_i is either the quantifier $\forall$ or the quantifier $\exists$ and the formula φ_0 is an open formula.
It is well known that every sentence is equivalent to a sentence written in prenex normal form.
A sentence is said to be universal if it is in prenex normal form and each quantifier is the universal quantifier $\forall$.

We then recall the notion of a structure in a signature Sig: A structure is in the form:

$$M = (|M|, (a^M)_{a \in Sig})$$

Where $|M|$ is a set called the universe of the structure, and for $a \in Sig$, a^M is the interpretation of a in M:
If f is a m-ary function symbol in Sig, then f^M is a function: $M^m \rightarrow M$.
If R is a m-ary relation symbol in Sig, then R^M is a relation: $R^M \subseteq M^m$.
If a is a constant symbol in Sig, then a^M is a distinguished element in M.

When M is a structure and φ is a sentence in the same signature Sig, we write $M \models \varphi$ for " M is a model of φ ", which means that φ is satisfied in the structure M. More details about these notions may be found in [CK73] or [Sch67].

When M is a structure in the signature Sig and Sig_1 is another signature such that $Sig_1 \subseteq Sig$, then the reduction of M to the signature Sig_1 is denoted

$M|Sig_1$. It is a structure in the signature Sig_1 which has same universe $|M|$ as M, and the same interpretations for symbols in Sig_1. Conversely an expansion of a structure M in the signature Sig_1 to a structure in the signature Sig has same universe as M and same interpretations for symbols in Sig_1.

When M is a structure in a signature Sig and $X \subseteq |M|$, we define:
$cl^1(X, M) = X \cup \bigcup_{\{f \text{ n-ary function of } Sig\}} f^M(X^n) \cup \bigcup_{\{a \text{ constant of } Sig\}} a^M$
$cl^{n+1}(X, M) = cl^1(cl^n(X, M), M)$ for an integer $n \geq 1$
and $cl(X, M) = \bigcup_{n\geq 1} cl^n(X, M)$ is the closure of X in M.

Let us now define locally finite sentences:

Definition 2.1 *A first order sentence φ (in the signature $S(\varphi)$= non logic symbols appearing in φ) is locally finite if and only if (iff):*

- *a) $M \models \varphi$ and $X \subseteq |M|$ imply $cl(X, M) \models \varphi$*
- *b) $\exists n \in \mathbb{N}$ such that $\forall M$, if $M \models \varphi$ and $X \subseteq |M|$, then $cl(X, M) = cl^n(X, M)$, where $cl(X, M)$ is the closure of X in M and $cl^n(X, M)$ is the subset obtained from X applying at most n times the functions of $S(\varphi)$. (closure in models of φ takes less than n steps).*

Notation. For a locally finite sentence φ, let n_φ be the smallest integer $n \geq 1$ verifying b) of the above definition.

Remark 2.2 Because of a) of Definition 2.1, a locally finite sentence φ is always equivalent to a universal sentence, So we may assume that we are still in that case.

Let us now introduce basic notations for words.
Let Σ be a finite alphabet whose elements are called letters. A finite word over Σ is a finite sequence of letters: $x = a_0...a_n$ where $\forall i \in [0; n]$ $a_i \in \Sigma$. We' ll denote $x(i) = a_i$ the $i + 1^{th}$ letter of x. The length of x is $|x| = n + 1$. The empty word will be denoted by λ and has 0 letter. Its length is 0. The set of finite words over Σ is denoted $\Sigma^\star$. $\Sigma^+ = \Sigma^\star - \{\lambda\}$ is the set of non empty words over Σ. A (finitary) language L over Σ is a subset of $\Sigma^\star$. Its complement (in $\Sigma^\star$) is ${}^cL = \Sigma^\star - L$. When $x = a_0...a_n$ is a word over Σ and the a_i are letters, we denote $x^R = a_n...a_0$ the word obtained from x by the reverse operation. The usual concatenation product of u and v will be denoted by $u.v$ or uv. $\mathbb{N}$ will be the set of non negative integers. For $V \subseteq \Sigma^\star$, we denote $V^\star = \{v_1...v_n/n \in \mathbb{N} \quad and \quad v_i \in V \quad \forall i \in [1; n]\}$.

A word over Σ may be considered as a structure in the following usual manner:

Let Σ be a finite alphabet. We denote P_a a unary predicate for each letter $a \in \Sigma$ and Λ_Σ the signature $\{<, (P_a)_{a\in\Sigma}\}$. Let σ be a word over the alphabet Σ, $|\sigma|$ is the length of the word σ. We may write that $|\sigma| = \{0, 1, ..., |\sigma|-1\}$. σ is identified to the structure $(|\sigma|, <^\sigma, (P_a^\sigma)_{a\in\Sigma})$ of signature Λ_Σ where $P_a^\sigma = \{i < |\sigma| /$ the $i+1^{th}$ letter of σ is an a $\}$.

Definition 2.3 *Let Σ be a finite alphabet and L be a language of finite words over the alphabet $\Sigma, L \subseteq \Sigma^\star$.*
Then L is a locally finite language $\longleftrightarrow$ there exists a locally finite sentence φ in a signature $\Lambda \supseteq \Lambda_\Sigma$ such that $\sigma \in L$ iff $\exists M, M \models \varphi$ and $M|\Lambda_\Sigma = \sigma$.
(where $M|\Lambda_\Sigma$ is the reduction of M to the signature Λ_Σ).
We then denote $L = L^\Sigma(\varphi)$, and to simplify, when there is no ambiguity, $L = L(\varphi)$ the locally finite language defined by φ.
The class of locally finite languages will be denoted LOC.

The empty word λ has 0 letter. It is represented by the empty structure.

Remark 2.4 The most natural convention concerning the possibility for a structure to have an empty domain is that an empty structure of signature Λ exists if Λ does not contain any constant symbol. A sentence of Λ is then true in the empty structure if under prenex form it begins by $\forall$ and false if it begins by $\exists$.
Let then $L(\varphi)$ be a locally finite language. If $\lambda \in L(\varphi)$, let $d \notin S(\varphi)$ and $\varphi' = \varphi \wedge (d = d)$, then φ' is locally finite, $S(\varphi') = S(\varphi) \cup \{d\}$ where d is a new constant symbol. And it holds that $L(\varphi') = L(\varphi) - \{\lambda\}$.
Conversely assume that $\lambda \notin L(\varphi)$. We then replace each constant symbol c of φ by a unary function symbol $c(u)$, each occurrence of c in φ by $c(u)$ for each constant c, we place $\forall u$ in front of φ and we add to φ the sentence $\forall uv \bigwedge_c c(u) = c(v)$.
The resulting sentence φ' is locally finite and verifies $L(\varphi') = L(\varphi) \cup \{\lambda\}$.

Remark 2.5 We call locally finite sentences the above defined sentences by analogy with the notion of a locally finite group. Each finite subset of a locally finite group generates a finite subgroup. Here each finite subset of a model of a locally finite sentence generates a finite submodel and moreover with a uniform upper bound over the iteration number which is necessary to obtain the generated submodel.

The notion of locally finite language seemingly looses its initial signification because a subword of a finite word always generates a finite word.
Nevertheless, there is moreover a uniform bound, and we may also consider words of infinite and even transfinite length of which each finite subword generates a finite subword. We' ll then keep the notion of language defined by a locally finite sentence.
The notion of locally finite language is very different of the usual notion of local language which represents a subclass of the class of rational languages. But whenever the things are well defined and precised, in a precise context and when there will be no ambiguity, we'll always call local languages the locally finite languages. So in the course of this paper.

2.2 First properties

Proposition 2.6 *a) The set of local sentences is recursively enumerable.*

b) It is undecidable to determine whether an arbitrary sentence φ is a local one.

Proof of a). Let $T_n(\varphi)$ be the set of terms obtained by applying at most n times the functions of $S(\varphi)$ and $T(\varphi) = \cup_{n\geq 1} T_n(\varphi)$.
We express by a first order formula the statement " $T_n(\varphi)$ generate $T(\varphi)$ " :

$$\bigwedge_{\{f function \in S(\varphi), t_1,\ldots,t_k \in T_n(\varphi)\}} \forall \bar{u} \bigvee_{t \in T_n(\varphi)} \exists (\text{sequence } \bar{v} \text{ of points of } \bar{u}) f(t_1(\bar{u}) \ldots t_k(\bar{u})) = t(\bar{v})$$

Then we enumerate all the proofs, checking whether they prove that:

$$\varphi \vdash [T_n(\varphi) \ \textit{generate} \ T(\varphi)], \quad \text{for some } n \in \mathbb{N}$$

.
If φ is local we obtain such a proof for some $n \geq n_\varphi$.

For an exact definition of a proof, see for example [CK73].

Proof of b). By Church's Theorem, it is undecidable to determine whether an arbitrary sentence φ is consistent, (recalling that a first order sentence is said to be consistent if it has at least one model, otherwise the sentence is said to be inconsistent). From this result we can deduce b): Indeed otherwise, admitting that a non consistent sentence is local, two cases may happen:
First case: φ is not local, then φ is consistent.
Second case: φ is local. As in the proof of a), we can find an integer $n \geq n_\varphi$

and therefore, if there exists a model of φ, we find one such model which cardinal is finite and bounded by an integer m obtained from n. Otherwise φ is not consistent.

In each case the algorithm determines whether φ is consistent, But " φ is consistent " is undecidable, then a contradiction would appear.

This negative result does not depend on the convention we have chosen:
" a non consistent sentence is local ". Indeed:

Proposition 2.7 *There does not exist any algorithm* **A** *which, for every sentence* ψ*, decide whether* ψ *is local* **and** *consistent.*

Proof. Otherwise let p be a propositional variable an let ψ_0 be a consistent local sentence . Using **A**, here is an algorithm which, for every sentence ψ, decide whether ψ is consistant.
We put $(\neg p \wedge \psi) \vee (\psi_0 \wedge p) = \theta$ as entry for A.

–If **A** answers that θ is local and consistent, we find $n \geq n_\theta$ as in the preceding proof, then we look for a model of ψ which cardinal is bounded by a bound depending on n. If we find such a model, then ψ is consistent, otherwise ψ is not.
– If **A** answers that θ is not " local and consistent ", then ψ is consistent (because θ being consistent as ψ_0 is, the answer of **A** means that ψ is not local, then ψ is consistent).

In a similar manner, we obtain:

Proposition 2.8 *There does not exist any algorithm* **A** *which, for every consistent sentence* ψ*, decide whether* ψ *is local.*

Per contra to these negative results, there exists a "recursive presentation" up to logical equivalence of all local sentences (two sentences φ and ψ in the same signature are said to be equivalent iff they have the same models, we then write $\varphi \equiv \psi$):

Theorem 2.9 *There exists a recursive set* **L** *of local sentences and a recursive function* **F** *such that:*

1) ψ *local* $\longleftrightarrow \exists \psi' \in L$ *such that* $\psi \equiv \psi'$.

2) $\psi' \in L \longrightarrow n_{\psi'} = F(\psi')$.

Proof. The elements of **L** are the $\psi \wedge C_n$, where ψ run over the universal formulas and C_n run over the universal formulas in the signature $S(\psi)$ which express that closure in a model takes at most n steps.
$\psi \wedge C_n$ is local and $n_{\psi\wedge C_n} \leq n$. Then we can compute $n_{\psi\wedge C_n}$, considering only finite models of cardinal $\leq m$, where m is an integer depending on n. And each local sentence ψ is equivalent to a universal formula θ, hence $\psi \equiv \theta \wedge C_{n_\psi}$.

These results are due to J-P. Ressayre.

From now on, in the sequel of this paper, we' ll assume that the local sentences, whenever they are not explicit, belong to this recursive set **L**.

2.3 Examples of locally finite languages

Remark that in the following examples, to prove that a language is local, we sometimes use the closure properties of the class *LOC* which will be shown in the next section.

Example 2.10 Rational languages. Recall that every rational language is a local language. This result is proved in [Res88] by induction on the complexity of a regular expression defining a rational language. It is clear that each language containing only one word of length 1, $a \in \Sigma$, is defined by the local sentence $\forall x[x = A \wedge P_a(A)]$ in the signature $\{<, P_a, A\}$, where A is a constant symbol. Then each rational language is local because *LOC* is closed under union, concatenation product and star operation (see Theorem 3.1 below).

Example 2.11 Dyck and Antidyck Languages.

Recall the following:

Definition 2.12 *The Antidyck language over two sorts of parentheses is the following language:* $Q_2'^{\star} = \{v \in (Y \cup \bar{Y})^\star / v \rightarrow^\star \lambda\}$, *where* $Y = \{y_1, y_2\}$, $\bar{Y} = \{\bar{y_1}, \bar{y_2}\}$ *and* $\rightarrow^\star$ *is the transitive closure of* $\rightarrow$ *defined in* $(Y \cup \bar{Y})^\star$ *by* $\forall y \in Y, y v_1 \bar{y} v_2 \rightarrow v_1 v_2$ *if and only if (iff)* $v_1 \in Y^\star$.

Remark 2.13 *This language may be seen as constituted by words with two sorts of parentheses, such that: " the first parenthesis to be opened is the first to be closed"*

Definition 2.14 *The Dyck language* $D_2'^{\star}$ *is defined in a similar manner but with* $v_1 y \bar{y} v_2 \rightarrow v_1 v_2$, $\forall v_1, v_2 \in (Y \cup \bar{Y})^\star$.

Remark 2.15 *This time we have: " the first opened parenthesis is the last to be closed".*

These definitions are generalized to Dyck and Antidyck languages $D_n'^{\star}$ and $Q_n'^{\star}$ with n sorts of parentheses , $n \geq 1$. Then we show the:

Proposition 2.16 *These languages $D_n'^{\star}$ and $Q_n'^{\star}$ are local.*

Proof. Let us show that $D_n'^{\star}$ is a local language. $D_n'^{\star}$ is a language over the alphabet $\Sigma = \{y_1, ..., y_n, \bar{y}_1, ..., \bar{y}_n\}$. The letters y_i

represent here the open parentheses, and the letters $\bar{y}_i$ represent the closed parentheses corresponding to the y_i.
We give a sentence φ_n in the signature $S(\varphi_n) = \{(P_a)_{a\in\Sigma}, <, s\}$, where s is a unary function symbol.
φ_n is the conjunction of the following sentences:

- $\forall xyz[(x \leq y \vee y \leq x) \wedge (x \leq y \wedge y \leq x \leftrightarrow x = y) \wedge (x \leq y \wedge y \leq z \rightarrow x \leq z)]$
  (this means: " $<$ is a linear order "),
- $\forall x[(\bigvee_{a\in\Sigma} P_a(x)) \wedge (\bigwedge_{(a,a')\in\Sigma^2, a\neq a'} \neg(P_a(x) \wedge P_{a'}(x)))]$
  (this means: $(P_a)_{a\in\Sigma}$ form a partition),
- $\forall x[P_{y_i}(x) \rightarrow x < s(x)]$, for each $i \in [1, n]$,
- $\forall xz[(P_{y_i}(x) \wedge P_{y_j}(z) \wedge x < z) \rightarrow s(z) < s(x)]$, for all $i, j \in [1, n]$,
- $\forall x[P_{y_i}(x) \leftrightarrow P_{\bar{y}_i}(s(x))]$, for each $i \in [1, n]$,
- $\forall x[s(s(x)) = x]$.

φ_n is equivalent to a universal formula and closure in its models takes only one step because $\forall x[s(s(x)) = x]$. Then φ_n is a local sentence and we easily verify that $L(\varphi_n) = D_n'^{\star}$.

Remark 2.17 *In the sequel, ($<$ is a linear order) and ($(P_a)_{a\in\Sigma}$ form a partition) are abreviations for the corresponding first order sentences written above.*

Example 2.18 $L = \{a^{n_1}ba^{n_2}b...a^{n_p}ba^{n_{p+1}}c...a^{n_{2p}}c / n_i \geq 0, p \geq 1\}$ **is local.**

Proof. This context free language is given by the local sentence φ which is the conjunction of:

- ($<$ is a linear order),
- (P_a, P_b, P_c, form a partition),
- $\forall xy[(P_b(x) \wedge P_c(y)) \rightarrow x < y]$,
- $\forall x[P_c(x) \leftrightarrow P_b(s(x))]$,
- $\forall x[s(s(x)) = x]$.
- $\forall x[P_a(x) \rightarrow s(x) = x]$,
- $\forall x[x \leq d \wedge P_c(d)]$.

φ is given in the signature $S(\varphi) = \{<, P_a, P_b, P_c, s, d\}$, where s is unary function symbol and d is a constant symbol.
φ is equivalent to a universal formula and closure in its models takes only two steps because it suffices to add the element which interprets d and then to take the closure by s in one step to obtain the closure in a model of φ.

Example 2.19 *The language* $L = \{0^n 1^p / 0 \leq n, p \leq 2^n\}$ *is local over the alphabet* $\Sigma = \{0, 1\}$*, while its complement* ${}^cL = \{u10v/u, v \in \Sigma^\star\} \cup \{0^n 1^p / 0 \leq n, p > 2^n\}$ *is not local.*

The language L is local: let φ be the following sentence which signature is $\{<, P_0, P_1, \in f\}$, where f is a binary function symbol and $\in$ is a binary relation symbol.
φ is the the conjunction of:

- ($<$ is a linear order),
- (P_0, P_1, form a partition),
- $\forall xy[(P_0(x) \wedge P_1(y)) \rightarrow x < y]$,
- $\forall xy[(x \in y) \rightarrow (P_0(x) \wedge P_1(y))]$,
- $\forall x[f(xx) \in x \leftrightarrow P_1(x)]$,
- $\forall xy[(x \neq y \wedge P_1(x) \wedge P_1(y)) \rightarrow ((f(xy) \in x \wedge f(xy) \notin y) \vee (f(xy) \in y \wedge f(xy) \notin x))]$,
- $\forall xy[P_0(x) \vee P_0(y) \rightarrow f(xy) = min(xy)]$.

A model M of φ is formed by two successive and disjoint segments P_0^M and P_1^M. One may consider that P_1^M is a subset of the set of nonempty subsets of P_0^M. In fact, there is an injection $j : P_1^M \to 2^{P_0^M} - \{$ emptyset $\}$, defined by $j(x) = \{z \in M / z \in^M x\}$.
This property in particular implies that card(P_1^M)$\leq 2^{card(P_0^M)}$. The function f^M defined by φ is then a choice function which permits to see that j is injective.
φ is equivalent to a universal formula and closure in its models takes only one step. And we easily see that $L = L(\varphi) \cup L(\varphi).1$ and so L is a local language. But if ${}^cL$ was local, for sufficiently large integer n, the word $0^n1^{2^n+1}$ of ${}^cL$ would have a strict subword containing 0^n and also being in ${}^cL$ (because the closure of 0^n in $0^n1^{2^n+1}$ would take at most n_φ steps, where n_φ does not depend on n).
This example is due to J-P. Ressayre.

Example 2.20 *Quasirational languages.*

The quasirational languages, also called non expansive, finite index, superlinear, have given rise to a great interest. With its two subfamilies of ultralinear and bounded languages, the family of quasirational languages has been more or less the subject of a large part of the work done about context free languages. The main advantage of these languages is the richness of their algebraic structures. Their various characterizations, using some very diverse concepts, like automata, grammars, closure under operators, algebraic expressions, provide efficacious investigation tools which generally fail in language theory. More, quasirational languages take a fundamental place into the context free languages general theory.
The family of quasirational languages is the closure under substitution of the cone of linear languages. It is also the family of languages which are generated by finite index context free grammars [Cre73].
Then we shall prove the following:

Theorem 2.21 *Every quasirational language is local.*

Proof. First state a lemma:

Lemma 2.22 *Every linear context free language is local.*

Proof. Let $L(G)$ be a linear context free language, over a finite alphabet Σ, which is generated by the linear grammar G of which production rules are: $A_i \to u_iB_iv_i$ for $1 \leq i \leq n$, and $C_i \to w_i$ for $1 \leq i \leq k$, where $\forall i$, $u_i, v_i, w_i \in \Sigma^\star$. The variables A_i, B_i, C_i not necessarily are distinct, but are

variables taken in a finite set given by G.
Let us now associate to $L(G)$ another linear language over the alphabet $\Gamma = \{c_1, ...c_n, d_1, ..., d_n, e_1, ..., e_k\}$. This new language is generated by the grammar G' of which the production rules are:
$A_i \to c_i B_i d_i$ for $1 \leq i \leq n$, and $C_i \to e_i$ for $1 \leq i \leq k$.
This language $L(G')$ is the set of words in the form $c_{i_1}...c_{i_j} e_i d_{i_j}...d_{i_1}$, where $1 \leq i \leq k$ and $i_1, ..., i_j$ are integers in $[1, n]$, $n \geq 1$.
The set $\{c_{i_1}...c_{i_j} e_i / c_{i_1}...c_{i_j} e_i d_{i_j}...d_{i_1} \in L(G')\}$ is a rational language, generated by the grammar of which production rules are $A_i \to c_i B_i$, $1 \leq i \leq n$, and $C_i \to e_i$, $1 \leq i \leq k$.
Therefore this language is local, defined by a local sentence φ. It is now easy to see that $L(G')$ is defined by a local sentence ψ, associating a word of $L(\psi)$ with each word of $L(\varphi)$:

Let $S(\psi) = \{P, a, s\} \cup S(\varphi) \cup \{P_{d_i} / 1 \leq i \leq n\}$, where P is a unary predicate symbol, a is a constant symbol and s is a unary function symbol.
Then ψ is the conjunction of the following sentences:

- ($<$ is a linear order),
- $((P_{c_i})_{1\leq i\leq n}, (P_{d_i})_{1\leq i\leq n}, (P_{e_i})_{1\leq i\leq k})$ form a partition,
- $\forall xy[P(x) \wedge \neg P(y) \to x < y]$,
- $\forall x_1...x_j \in P[\varphi_0(x_1, ..., x_j)]$, where $\varphi = \forall x_1...x_j \varphi_0(x_1, ..., x_j)$ with φ_0 an open formula,
- $\forall x_1...x_m \in P[f(x_1, ..., x_m) \in P]$, for each m-ary function f of $S(\varphi)$,
- $\forall x_1...x_m[\bigvee_{1\leq i\leq m} \neg P(x_i) \to f(x_1, ..., x_m) = min(x_1, ..., x_m)]$, for each m-ary function f of $S(\varphi)$,
- $P(c)$, for each constant c of $S(\varphi)$,
- $P(a) \wedge \forall x(P(x) \to x \leq a)$,
- $s(a) = a$,
- $\forall x \in P[x \neq a \to \neg P(s(x))]$,
- $\forall xy[P(x) \wedge P(y) \wedge x < y \to s(y) < s(x)]$,
- $\forall x[s(s(x)) = x]$,
- $\forall x[P(x) \wedge P_{c_i}(x) \leftrightarrow P_{d_i}(s(x))]$, for each integer $i \in [1, n]$.

ψ is equivalent to a universal formula and closure in its models takes at most $n_\varphi + 2$ steps: one takes closure under s then by the functions of $S(\varphi)$, and then again by s. One can check that $L(\psi) = L(G')$ holds by construction.
Let now the morphism $h : \Gamma^\star \to \Sigma^\star$ defined by $h(c_i) = u_i$, $h(d_i) = v_i$, for $1 \leq i \leq n$, and $h(e_i) = w_i$ for $1 \leq i \leq k$. Then $h(L(G')) = L(G)$ hence $L(G)$ is the image of the local language $L(\psi)$ by the morphism h and, using the fact that LOC is closed under morphism (proved in next section), we can infer that $L(G)$ is local.
Then each linear context free language is local.

To end the proof of the Theorem, recall that the family of quasirational languages is the closure under substitution of the family of linear languages. LOC being closed under substitution, (result proved in next section), the preceding lemma implies that quasirational languages are local.

Remark that context free languages L such that each word in L is in the form $xca^{|x|}$, where $a \in \Sigma$, $c \notin \Sigma$, Σ is a finite alphabet and $x \in \Sigma^\star$, are important languages for the syntax of many programming languages. The set $\{x/xca^{|x|} \in L\}$ is in that case rational hence local, and this permits to easily prove that L is local.
In a similar manner, the context free languages of which each word is in the form xcx^R, important for programming, are local.

Example 2.23 Denote CF the class of context free languages, and OC the class of one counter languages. Then $RAT \subsetneq OC \subsetneq CF$ [Ber79].
$RAT \subseteq LOC$.
$\{a^n b^n / n \geq 1\}$ is in LOC and in $OC - RAT$.
The Dyck language with two sorts of parentheses is in $CF - OC$ and in LOC.

Then many context free languages are in LOC, but it is an open question to know whether every context free language is local.

Example 2.24 The cone of Greibach languages is the closure under substitution of the family $LIN \cup OC$, where LIN is the family of linear languages and OC is the family of one counter languages. We have seen that there are local languages which are generators of the family of context free languages, like the Dyck languages $D_n'^\star$, for $n \geq 2$.
There are local languages in $NGE - GRE$, the family of context free languages which are neither generator nor Greibach languages.
An example is given by the following language:

Let $\hat{Z}_n = \{z_1, ..., z_n, \bar{z_1}, ..., \bar{z_n}\} = Z_n \cup \bar{Z}_n$
Where $Z_n = \{z_1, ..., z_n\}$ and $\bar{Z}_n = \{\bar{z}_1, ..., \bar{z}_n\}$.
Let us define the substitution s over $\hat{Z}_n$ by:
$s(z_i) = z_i$ and $s(\bar{z}_i) = \bar{z}_i \bar{Z}_n^\star$ for $1 \leq i \leq n$.

The image δ_n of the Dyck language over the alphabet $\hat{Z}_n$ (with n sorts of parentheses) by the substitution s is then context free and, for $n \geq 2$, $\delta_n \in NGE - GRE$ holds.
This language is local, because the Dyck languages and the languages $\bar{z}_i \bar{Z}_n^\star$ are local, and LOC is closed under substitution by the Theorem 3.3 of next section.

The question is open to know whether $OC \subset LOC$, which would imply that $GRE \subset LOC$.

Example 2.25 The Hardest Context Free Language or the non deterministic version of the Dyck language.

This language is defined over the alphabet $\Sigma_n = \hat{Z}_n \cup \{[,], +\}$; A block is a word in the form $[u_1 + u_2 + ... + u_p]$; So the set of blocks is the rational language $[(\hat{Z}_n \cup \{+\})^\star]$.
A factor u in a block (maximal factor over $\hat{Z}_n$) is called a choice in this block. Then $f \in H_n$ iff $f = f_1 f_2 ... f_r$ where: each f_i is a block and there exists a choice u_i in every block f_i such that $u_1 u_2 ... u_r \in D_n'^\star$.

All these languages H_n are context free and, for $n \geq 2$, H_n is a generator of the rational cone CF. These languages have been studied because they generate the family CF without using direct morphisms. Indeed Shamir-Greibach Theorem states that:
A language L is context free iff there exists a morphism h such that:
$L - \{\lambda\} = h^{-1}(H_2)$.
Let us show that H_n is local.
Consider first the sentence ψ_n which is the conjunction of:

- ($<$ is a linear order),
- ($(P_a)_{a \in \Sigma_n}$ form a partition),
- $\forall xy(P_[(x) \rightarrow x \leq y)$,
- $\forall xy(P_](x) \rightarrow y \leq x)$,
- $\forall x(P(x) \rightarrow \bigvee_{a \in \hat{Z}_n} P_a(x))$,

- $\forall x(P(x) \leftrightarrow b_1 < x < b_2)$,
- $P_{[}(b_1) \vee P_{+}(b_1)$,
- $P_{[}(a_1) \wedge P_{]}(a_2)$,
- $P_{]}(b_2) \vee P_{+}(b_2)$,

The signature of ψ_n is $S(\psi_n) = \{(P_a)_{a\in\Sigma_n}, <, P, b_1, b_2, a_1, a_2\}$, where P is a unary predicate symbol and b_1, b_2, a_1, a_2 are constant symbols.
A model M of ψ_n provides a block $M|\Lambda_{\Sigma_n}$ and $P^M|\Lambda_{\Sigma_n}$ gives a choice in this block.

Let then the signature $S(\varphi_n) = \{(P_a)_{a\in\Sigma_n}, <, P, I, s, b_1, b_2, a_1, a_2\}$, where P is a unary predicate symbol and I, s, b_1, b_2, a_1, a_2 are unary function symbols.
Let $\varphi = \forall xyz\varphi_0(xyz)$ a local sentence defining $D_n'^{\star}$ with φ_0 an open formula, $S(\varphi) = \{(P_a)_{a\in\hat{Z}_n}, < s\}$, s a unary function. (See Example 2.11)
Now define :

$$\varphi_n = \psi_n^{\star} \wedge [\forall xyz \in P \quad \varphi_0(xyz)] \wedge \forall x(\neg P(x) \to s(x) = x),$$

where $\psi_n^{\star}$ is the conjunction of:

- ($<$ is a linear order),
- ($(P_a)_{a\in\Sigma_n}$ form a partition),
- $\forall xy[(I(y) \leq y) \wedge (y \leq x \to I(y) \leq I(x)) \wedge (I(y) \leq x \leq y \to I(x) = I(y))]$,
- $\forall xy[I(x) = I(y) \to e(x) = e(y)]$, for $e \in \{b_1, b_2, a_1, a_2\}$,
- $\forall x\psi_n^x$,
  where ψ_n^x is the local sentence ψ_n in which the constants b_1, b_2, a_1, a_2 are replaced by the terms $b_1(x), b_2(x), a_1(x), a_2(x)$ and each quantifier is relativized to the set $\{y/I(y) = I(x)\}$.

φ_n is equivalent to a universal formula and closure in each of its models takes at most two steps: (take first closure under s, then by b_1, b_2, a_1, a_2 and I.
Hence φ_n is local and, by construction $L^{\Sigma_n}(\varphi_n) = H_n$ holds.
We'll see in next section that LOC is not closed under inverse morphism; Then we cannot use this result in order to prove, by Shamir-Greibach Theorem, that $CF \subset LOC$.

Example 2.26 Languages of bigger complexity than that of context free languages.

Let $NEST$ be the family of languages which are recognized by a non erasing stack automata, ST be the family of languages which are recognized by a stack automata, and CS be the family of context sensitive languages which are generated by a context sensitive grammar.
It holds that: $CF \subsetneq NEST \subsetneq ST \subsetneq CS$. [Ogd69].
There are local languages in each family of this hierarchy.

$\{a^{n^2}b^n/n \geq 1\}$ is in $NEST - CF$.
$\{b^n a^{n^2}/n \geq 1\}$ is in $ST - NEST$.
$\{a^{n^2}b^{n^2}c^{n^2}/n \geq 1\}$ is in $CS - ST$. [Ogd69].

Proof. We shall prove that the language $\{b^n a^{n^2}/n \geq 1\}$ is local; some analogous methods proving that $\{a^{n^2}b^n/n \geq 1\}$ and $\{a^{n^2}b^{n^2}c^{n^2}/n \geq 1\}$ are local.
Let $S(\psi) = \{P_a, P_b, <, d, f, p_1, p_2\}$ where f is a 2-ary function symbol, p_1 and p_2 are unary function symbols and d is a constant symbol.
The sentence ψ is the conjunction of:

- ($<$ is a linear order),
- (P_a, P_b form a partition),
- $\forall xy[(P_b(x) \wedge P_a(y)) \rightarrow x < y]$,
- $\forall xyzt[(P_b(x) \wedge P_b(y) \wedge P_b(z) \wedge P_b(t) \wedge (x \neq z \vee y \neq t)) \rightarrow (f(xy) \neq f(zt) \wedge P_a(f(xy)))]$,
- $\forall xy[(P_a(x) \vee P_a(y)) \rightarrow f(xy) = x]$,
- $\forall x[P_a(x) \rightarrow P_b(p_1(x)) \wedge P_b(p_2(x)) \wedge f(p_1(x)p_2(x)) = x]$,
- $\forall x[P_b(x) \rightarrow p_1(x) = p_2(x) = x]$,
- $\forall x[x \leq d]$.

ψ is equivalent to a universal formula and closure in each of its models takes at most three steps: (one first takes first closure under d, then under p_1 and p_2, and then under f).
Hence ψ is local and $L(\psi) = \{b^n a^{n^2}/n \geq 1\}$ holds by construction.

Example 2.27 In comparison with Kasai Hierarchy

We investigate here another hierarchy of families of languages which are located betweeen CF and CS, introduced by Takumi Kasai, [Kas70].

Many programming languages cannot be represented by context free languages, hence he studied grammars which are more powerful than the context free ones: the state grammars. Making restrictions over these grammars, Kasai obtained an infinity of AFL (Abstract Family of Languages), families of state languages which are closed under rational operations $\cup, ., +$, and under rational transductions.

Definition 2.28 ([Kas70]) *A state grammar is a sextuple* $G = (K, V, \Sigma, P, p_0, \sigma)$ *where:*
(1) K *is a non empty finite set (of states).*
(2) V *is a finite set of symbols and* $\Sigma \subseteq V$.
(3) σ *is an element of* $V - \Sigma$.
(4) p_0 *is an element of* K.
(5) P *is a finite subset of* $K \times (V - \Sigma) \times K \times V^+$ *where* $V^+ = \cup_{i=1}^{\infty} V^i$.

An element (p, ξ, q, u) of P is called a production and is usually written $(p, \xi) \to (q, u)$. $V - \Sigma$ is the set of variables and a variable ξ is said to be applicable under a state p if $(p, \xi) \to (q, u)$ is in P for some $q \in K$ and u in V^+. Let $G = (K, V, \Sigma, P, p_0, \sigma)$ be a state grammar, and let $\Rightarrow$ be a relation on $K \times V^+$ defined as follows:
Let $p \in K$ and $w = x\xi y \in V^+$. If this ξ is the leftmost occurrence of applicable variables in w under p and $(p, \xi) \to (q, u)$ is in P, then we write $(p, x\xi y) \Rightarrow (q, xuy)$.
If this ξ is the j-th variable in w, then we sometimes write $\stackrel{j}{\Rightarrow}$ instead of $\Rightarrow$.
For α and β in $K \times V^+$, write $\alpha \stackrel{\star}{\Rightarrow} \beta$ if either $\alpha = \beta$ or there exist $\alpha_0, ..., \alpha_r$ such that $\alpha_0 = \alpha$, $\alpha_r = \beta$, and $\alpha_i \Rightarrow \alpha_{i+1}$ for each i. The sequence $\alpha_0, ..., \alpha_r$ is called a derivation (of length r) and is denoted $\alpha_0 \Rightarrow ... \Rightarrow \alpha_r$.
The subset of Σ^+, $L(G) = \{w \in \Sigma^+ / (p_0, \sigma) \stackrel{\star}{\Rightarrow} (q, w)$ for some $q \in K\}$ is called a state language, generated by the grammar G.

Definition 2.29 ([Kas70]) *Let* $G = (K, V, \Sigma, P, p_0, \sigma)$ *be a state grammar and let* n *be a positive integer. A* n*-limited derivation is a derivation:*
$\alpha_0 \stackrel{j(1)}{\Rightarrow} \alpha_1 \stackrel{j(2)}{\Rightarrow} ... \stackrel{j(r)}{\Rightarrow} \alpha_r$ *such that* $j(i) \leq n$ *for each* i.
In this case we sometimes write $\alpha_0 \stackrel{n^\star}{\Rightarrow} \alpha_r$ *instead of* $\alpha_0 \stackrel{\star}{\Rightarrow} \alpha_r$, *in order to indicate that it is realized by a* n*-limited derivation.*
Then $L(G, n) = \{w \in \Sigma^+ / (p_0, \sigma) \stackrel{n^\star}{\Rightarrow} (q, w)$ *for some* $q \in K\}$
G *is said to be a state grammar of degree* n *iff* $L(G) = L(G, n)$.

A state language is said to be of degree n iff it is generated by a state grammar of degree n. Otherwise this state language is of infinite degree.

L_1 is the family of context free languages which do not contain the empty word.
L_n is the family of state languages of degree n.
$L_\infty = \cup_{n\geq 1} L_n$ is the family of state languages of finite degree.
L_ω is the family of state languages which is equal to the family of context sensitive languages.
T. Kasai proved that all the following inclusions are strict:
$L_1 \subsetneq L_2 \subsetneq ... \subsetneq L_n \subsetneq L_{n+1} \subsetneq ... \subsetneq L_\infty \subsetneq L_\omega$.
Let us cite a property of finite degree state languages:
If $L \in L_n$ for some $n \geq 1$, then $\{|w|/w \in L\}$ is an ultimately periodic set of integers. In comparison with local languages, it holds that:

Proposition 2.30 *There exist local languages at each level of this hierarchy; more precisely, $\forall n \geq 1$, there exists a local language $L(\varphi_n)$ in $L_{n+1} - L_n$ and there exists a local language in $L_\omega - L_\infty$ i.e. a (context sensitive) state language of infinite degree.*

Proof. The language $\{a_1^k a_2^k ... a_{4n+2}^k / k \geq 1\}$ over the alphabet $\Sigma_n = \{a_1, ..., a_{4n+2}\}$ is in $L_{n+1} - L_n$, for $n \geq 1$.
And the language $\{b^n a^{n^2} / n \geq 1\}$ over the alphabet $\Sigma = \{a, b\}$ is in L_ω but not in L_∞ ,[Kas70].
Now it is easy to check that these languages are local.

Example 2.31 There exist NP-complete languages in LOC.

This result is stated in [Res88].
For example the NP-complete language $CLIQUE$ is a local language:
For a finite graph G and an integer $n \geq 1$, $(G, n) \in CLIQUE$ iff G contains a clique of cardinal n (i.e. a set C of vertices of cardinal n such that every edge between two vertices of C is in G).

A graph $G = (V, E)$ is defined by the set V of its vertices and the set $E \subseteq V \times V$ of its edges, with $\forall (u, v) \in V^2 [(u, v) \in E \rightarrow (v, u) \in E]$ and $[(u, u) \notin E]$.
We code a word in $CLIQUE$ in the following manner:
$card(V)$ zeros followed by an a followed by $(cardV)^2$ letters in $\{0, 1\}$ coding E followed by a b followed by n times the letter 1.
If $V = \{v_1, ..., v_p\}$, we obtain the following word:

$0^p a[v_1v_1][v_1v_2]...[v_1v_p][v_2v_1][v_2v_2]...[v_2v_p]...[v_pv_1][v_pv_2]...[v_pv_p]b1^n$
where $[v_iv_j] = 0$ if $(v_i, v_j) \in E$ and $[v_iv_j] = 1$ if $(v_i, v_j) \notin E$.

Let then φ the following sentence, conjunction of:

- ($<$ is a linear order),
- (P_0, P_1, P_a, P_b) form a partition,
- (P, Q, R, P_a, P_b) form a partition,
- $\forall xyz[P(x) \wedge Q(y) \wedge R(z) \to x < A < y < B < z]$,
- $P_a(A) \wedge \forall x(P_a(x) \leftrightarrow x = A)$,
- $P_b(B) \wedge \forall x(P_b(x) \leftrightarrow x = B)$,
- $\forall xyz \in P[y < z \to f(xy) < f(xz)]$,
- $\forall xyzt \in P[y < z \to f(yt) < f(zx)]$,
- $\forall xy \in P[Q(f(xy))]$,
- $p_1(A) = p_2(A) = A$,
- $p_1(B) = p_2(B) = B$,
- $\forall xy[\neg P(x) \vee \neg P(y) \to f(xy) = min(xy)]$,
- $\forall x[P(x) \vee R(x) \to p_1(x) = p_2(x) = x]$,
- $\forall x[Q(x) \to f(p_1(x)p_2(x)) = x]$,
- $\forall x[Q(x) \to P(p_1(x)) \wedge P(p_2(x))]$,
- $\forall x[P(x) \to P_0(x)]$,
- $\forall x \in Q[P_0(x) \vee P_1(x)]$,
- $\forall x \in R[P_1(x)]$,
- $\forall x[\neg R(x) \to i(x) = x]$,
- $\forall x[R(x) \to P(i(x))]$,
- $\forall xy[R(x) \wedge R(y) \wedge x \neq y \to i(x) \neq i(y)]$,
- $\forall xy[R(x) \wedge R(y) \wedge x \neq y \to P_0(f(i(x)i(y)))]$,

- $\forall xy[P(x) \wedge P(y) \rightarrow (P_0(f(xy)) \leftrightarrow P_0(f(yx)))]$,
- $\forall x \in P[P_1(f(xx))]$.

The signature of φ is $S(\varphi) = \{<, P_0, P_1, P_a, P_b, P, Q, R, f, A, B, i, p_1, p_2\}$, where P, Q, R are unary predicate symbols, A, B are constant symbols, f is a binary function symbol and i, p_1, p_2 are unary function symbols.
φ is equivalent to a universal formula and closure in each of its models takes at most two steps: one takes the closure under A, B, p_1, p_2, i and then under f. Hence φ is a local sentence .
We then check that $L(\varphi)^{\{a,b,0,1\}} = CLIQUE$.

Another example of NP-complete language: The language which is connected with the problem VC: " Vertex Cover ".
Let $G = (V, E)$ be a graph and let k be a non negative integer $\leq card(V)$.
$(G, k) \in VC \leftrightarrow \exists V' \subseteq V$ such that $card(V') \leq k$ and V' covers the graph, i.e. for each edge $(u, v) \in E$, $\{u, v\} \cap V'$ is not empty.

We code a word of VC in the same manner as for $CLIQUE$.
Let then $S(\psi) = \{<, P_0, P_1, P_a, P_b, P, Q, R, V', f, A, B, s, p_1, p_2\}$, where V', P, Q, R are unary predicate symbols, A, B are constant symbols, f is a binary function symbol and s, p_1, p_2 are unary function symbols.

Let then ψ the following sentence, conjunction of:

- ($<$ is a linear order),
- (P_0, P_1, P_a, P_b) form a partition,
- (P, Q, R, P_a, P_b) form a partition,
- $\forall x[P_a(x) \leftrightarrow x = A]$,
- $\forall x[P_b(x) \leftrightarrow x = B]$,
- $\forall xyz[P(x) \wedge Q(y) \wedge R(z) \rightarrow x < A < y < B < z]$,
- $\forall xyz \in P[y < z \rightarrow f(xy) < f(xz)]$,
- $\forall xyzt \in P[y < z \rightarrow f(yt) < f(zx)]$,
- $\forall xy \in P[Q(f(xy))]$,
- $\forall x[Q(x) \rightarrow f(p_1(x)p_2(x)) = x]$,

- $\forall x[Q(x) \rightarrow P(p_1(x)) \wedge P(p_2(x))]$,
- $\forall x[\neg Q(x) \rightarrow p_1(x) = p_2(x) = x]$,
- $\forall x[P(x) \rightarrow P_0(x)]$,
- $\forall x[Q(x) \rightarrow P_0(x) \vee P_1(x)]$,
- $\forall x[R(x) \rightarrow P_1(x)]$,
- $\forall xy[P(x) \wedge P(y) \rightarrow (P_0(f(xy)) \leftrightarrow P_0(f(yx)))]$,
- $\forall x \in P[P_1(f(xx))]$.
- $\forall x[V'(x) \rightarrow P(x)]$,
- $\forall x[V'(x) \rightarrow R(s(x)) \wedge s(s(x)) = x]$,
- $\forall x[R(x) \rightarrow P(s(x))]$,
- $\forall xy[R(x) \wedge R(y) \wedge x \neq y \rightarrow s(x) \neq s(y)]$,
- $\forall x[\neg R(x) \wedge \neg V'(x) \rightarrow s(x) = x]$,
- $\forall xy[P(x) \wedge P(y) \wedge P_0(f(xy)) \rightarrow V'(x) \vee V'(y)]$.

ψ is equivalent to a universal formula and closure in each of its models takes at most three steps: one takes the closure under A, B, p_1, p_2 and then under s, then under f. Hence ψ is a local sentence .
By construction we can check that in a model M of ψ, V'^M represents a subset of P^M of cardinal $\leq k \leq card(P^M)$, $k = card(R^M)$, which covers the graph. Then it holds that $L(\psi)^{\{a,b,0,1\}} = VC$.

Example 2.32 There exists a set of integers in **P** which is not the spectrum of any local language, as $\{2^n/n \geq 1\}$. (Result of J-P. Ressayre)

Recall that the spectrum of a first order sentence φ is the subset of $\mathbb{N}$ defined by: $Sp(\varphi) = \{n \in \mathbb{N} \ / \ \exists M \models \varphi$ such that $card(M) = n\}$.

Example 2.33 By methods which are similar to those used in example 2.26, we easily show that $p(\mathbb{N})$ is a spectrum of a local language for every polynomial p with coefficients in $\mathbb{N}$. And more generally if A is the spectrum of a local language, then $p(A)$ is the spectrum of a local language, for every polynomial p with coefficients in $\mathbb{N}$.

Let us show that $L = \{a^n / n \text{ is a non prime integer} \geq 0\}$ is a local language.

Let $S(\varphi) = \{<, P_a, P, b, c, f, p_1, p_2\}$, where P is a unary predicate symbol, b, c are constant symbols, f is a binary function symbol and p_1, p_2 are unary function symbols.

Let then φ the following sentence, conjunction of:

- ($<$ is a linear order),
- $\forall x P_a(x)$,
- $\forall xy[P(x) \wedge \neg P(y) \rightarrow x < y]$,
- $P(b) \wedge P(c) \wedge c < b$,
- $\forall xy[P(x) \wedge P(y) \wedge y < b \rightarrow \neg P(f(xy))]$,
- $\forall xyzt[P(x) \wedge P(y) \wedge P(z) \wedge P(t) \wedge y < b \wedge t < b \wedge (x \neq z \vee y \neq t) \rightarrow f(xy) \neq f(zt)]$,
- $\forall xy[\neg P(x) \vee b \leq y \rightarrow f(xy) = x]$,
- $\forall x[P(x) \rightarrow p_1(x) = p_2(x) = x]$,
- $\forall x[\neg P(x) \rightarrow f(p_1(x)p_2(x)) = x \wedge p_2(x) < b \wedge P(p_1(x))]$.

We then check that this sentence φ is local because it is equivalent to a universal formula and closure in each of its models takes at most two steps: one takes the closure under p_1, p_2 and then under f.
If $M \models \varphi$ and $card(P^M) = m$, then $card(M) = m + nm$ with $1 \leq n < m$. Hence $card(M) = m(n+1)$ with $2 \leq n+1 \leq m$. Conversely we check that for every non prime integer in the form $m(n+1)$ with $2 \leq n+1 \leq m$, there exists a model of φ of cardinal $m(n+1)$. Therefore the spectrum of φ is $\{n \geq 4 \quad /n \text{ is a non prime integer }\}$. ¿From what we easily deduce that the language $L = \{a^n \ / \ n \text{ is a non prime integer} \geq 0\} = \{\lambda, a\} \cup L(\varphi)$ is a local language.
So the spectrum $\{n \in \mathbb{N} \quad /n \text{ is non prime }\}$ is the spectrum of a local language.
Its complement, $\{n \in \mathbb{N} \quad /n \text{ is prime }\}$ is not the spectrum of a local language, because the spectrum of an infinite local language always contains a set in the form $p(\mathbb{N})$, where p is a polynomial with positive integers as coefficients, [Fin99]. Indeed a result of number theory states that a polynomial with coefficients in $\mathbb{N}$ cannot take only prime number values over $\mathbb{N}$.

Remark that the question to know whether the complement of a spectrum is a spectrum is connected with the algorithmic complexity problem $\mathbf{NP} = \mathbf{co} - \mathbf{NP}$. This problem has a negative answer when it is restricted to spectra of local sentences. See [Fag93] for more results and references about spectra of first order sentences.

3 Closure properties

Recall the following:

Theorem 3.1 ([Res88]) *Loc is closed under union, catenation product, and operation* $\star$.

Proof. the proof of closure under union and catenation product is easy.
To prove closure under star operation, J-P. Ressayre defined a new operation over local sentences:
For each local sentence φ, $\varphi^\star$ is a local sentence in the signature $\mathrm{S}(\varphi^\star)$ which is the signature of φ to which is added a unary function symbol I and in which every constant symbol e is replaced by a unary function symbol $e(x)$.

For a local sentence φ, $\varphi^\star$ is the sentence defined by the conjunction of:

- ($<$ is a linear order),
- $\forall yz[I(y) \leq y$ and $(y \leq z \rightarrow I(y) \leq I(z))$ and $(I(y) \leq z \leq y \rightarrow I(z) = I(y))]$,
- $\forall xy[I(x) = I(y) \rightarrow e(x) = e(y)]$, for each constant e of the signature $\mathrm{S}(\varphi)$ of φ,
- $\forall x_1 ... x_n[(\bigvee_{i,j \leq n} I(x_i) \neq I(x_j)) \rightarrow f(x_1...x_n) = min(x_1...x_n)]$, for each n-ary function f of $\mathrm{S}(\varphi)$,
- $\forall x_1 ... x_n[(\bigwedge_{i,j \leq n} I(x_i) = I(x_j)) \rightarrow I(f(x_1...x_n)) = I(x_1)]$, for each n-ary function f of $\mathrm{S}(\varphi)$,
- $\forall x \varphi^x$, where φ^x is the local sentence φ in which every constant e is replaced by the term e(x) and each quantifier is relativized to the set $\{y \setminus I(y) = I(x)\}$.

Remark 3.2 *The models of* $\varphi^\star$ *are essentially direct sums of models of* φ, *and* $L^\Sigma(\varphi^\star) = (L^\Sigma(\varphi))^\star$.

We shall prove the following:

Theorem 3.3 *The class LOC of local languages is closed under substitution, morphism, inverse alphabetic morphism. LOC is not closed under intersection, intersection with a rational language, complementation, inverse morphism.*
LOC is neither closed under quotient, nor under quotient by a context free language, but LOC is closed under quotient by a finite language.

Proof.

a) Closure under substitution.
Let then $\Sigma = \{a_1, ..., a_n\}$ be a finite alphabet and let f be a substitution: $\Sigma \rightarrow P(\Gamma^\star)$, $a_i \rightarrow L_i$ where $\forall i \in [1; n]$, L_i is a local language defined by the sentence φ_i, over the alphabet Γ. More assume that the empty word is not in L_i. We may also assume that the signatures of the sentences φ_i verify $S(\varphi_i) \cap S(\varphi_j) = \{<, (P_a)_{a\in\Gamma}\}$ for $i \neq j$. Let now $L \subseteq \Sigma^\star$ be a local language defined by a local sentence φ. We'll denote by Q_{a_i} the unary predicate of $S(\varphi)$ which indicates the places of the letters a_i in a word of L, so as that if $a_i \in \Gamma \cap \Sigma$ for an indice i, there will be two distinct predicates Q_{a_i} and P_{a_i}. We' ll also assume, possibly differently naming function, constant, and other predicate symbols of $S(\varphi)$, that $\forall i \in [1, ..., n]$, $S(\varphi_i) \cap S(\varphi) = \{<\}$. Then we now construct a local sentence ψ which defines the language $f(L)$: ψ is the conjunction of the following sentences, which meaning is explained below:

- " $<$ is a linear order ",
- $\forall xy[(I(y) \leq y) \wedge (y \leq x \rightarrow I(y) \leq I(x)) \wedge (I(y) \leq x \leq y \rightarrow I(x) = I(y))]$,
- $\forall x[I(x) = x \leftrightarrow P(x)]$,
- $P(c)$, for each constant c of $S(\varphi)$,
- $\forall x_1...x_k[R(x_1...x_k) \rightarrow P(x_1) \wedge .... \wedge P(x_k)]$, for each predicate $R(x_1...x_k)$ of $S(\varphi)$,
- $\forall x_1...x_j[(P(x_1) \wedge ... \wedge P(x_j)) \rightarrow P(f(x_1...x_j))]$, for each j-ary function symbol f of $S(\varphi)$,
- $\forall x_1...x_j[(\bigvee_{1\leq i\leq j} \neg P(x_i)) \rightarrow f(x_1...x_j) = min(x_1...x_j)]$, for each j-ary function symbol f of $S(\varphi)$,

- $\forall x_1...x_m[(P(x_1)\wedge...\wedge P(x_m)) \to \varphi_0(x_1...x_m)]$, where $\varphi = \forall x_1...x_m\varphi_0(x_1...x_m)$ with φ_0 an open formula,

- $\forall x_1...x_j[\bigvee_{i,k\leq j}(I(x_i) \neq I(x_k)) \to f(x_1...x_j) = min(x_1...x_j)]$, for every function f of $S(\varphi_l)$ for an integer $l \leq n$,

- $\forall xy_1...y_j[(\bigwedge_{1\leq l\leq j} I(y_l) = I(x)) \to I(f(y_1...y_j)) = I(x)]$, for each j-ary function symbol f of $S(\varphi_i)$ for an integer $i \leq n$,

  Finally, for each $i \leq n$:

- $\forall xy_1...y_p[(\bigwedge_{1\leq l\leq p} I(y_l) = I(x) \wedge Q_{a_i}(I(x))) \to \varphi_i^0(y_1...y_p) \wedge \{(e_j(y_1) = I(x)\wedge f_j(y_1...y_p) = y_1\wedge\neg R_j(y_1...y_p)\wedge\bigwedge_{e_i\in S(\varphi_i)} e_i(y_1) = e_i(x)$; where $n \geq j \neq i$, and e_j, f_j, R_j run over the constants, functions, and predicates of $S(\varphi_j)\}]$,

Above , 1) to each constant e_l of $S(\varphi_l)$ is associated a new unary function $e_l(y)$ and 2) whenever $\varphi_i = \forall y_1...y_p\psi_i(y_1...y_p)$ with ψ_i an open formula, φ_i^0 is ψ_i in which every constant e_i has been replaced by the function $e_i(y)$.

Construction of ψ :
Using the function I which marks the first letters of the subwords, we divide a word into subwords. In every model M of ψ, the set of the "first letters of subwords" , P^M, grows richer in a model of φ (therefore will constitute a word of L).
Then we "substitute" : for each letter a_i in P^M, we substitute a word of L_i, using for that the formula φ_i.
Then if closure takes at most $m(\varphi)$ (respectively $m(\varphi_i)$) steps in every model of φ (respectively of φ_i), then closure takes at most $[m(\varphi) + 2sup_i(m(\varphi_i))]$ steps in each model of ψ.
Therefore ψ is a local sentence and by construction ψ defines the language $f(L)$.

When the languages L_i may contain the empty word λ,
consider a substitution f: $\Sigma \to P(\Gamma^\star)$, $a_i \to L_i$ as above and let $c \notin \Sigma \cup \Gamma$. Define f^c by: $f^c : \Sigma \to P((\Gamma\cup\{c\})^\star)$, $a_i \to L_i$ if $\lambda \notin L_i$ and $a_i \to L_i - \{\lambda\}\cup\{c\}$ otherwise. By the preceding proof, the language $f^c(L)$ is local. The local sentence defining $f^c(L)$ contains the unary predicate P_c which marks the places of the letters c in every word. The "projection" of the models of this sentence on the predicate $\neg P_c$ gives the local language $f(L)$. Indeed according to a result of J-P. Ressayre: if $L(\varphi)$ is a local language over an

alphabet Σ and P is a unary predicate of $S(\varphi)$, then $\{\sigma \in \Sigma^\star / \exists M \models \varphi$ and $M|P^M|\Lambda_\Sigma = \sigma\}$ is a local language.

b) Closure under morphism.
This is a particular case of the preceding one.

b) Closure under inverse alphabetic morphism.
Let f be an alphabetic morphism: $\Sigma \to \Gamma \cup \{\lambda\}$, where Σ and Γ are finite alphabets.
Let $\Sigma' = \{a \in \Sigma / f(a) = \lambda\}$.
Let $L(\varphi) \subseteq \Gamma^\star$ be a local language.
We first replace in φ the letter predicates $(P_a)_{a\in\Gamma}$ by $(Q_a)_{a\in\Gamma}$.
Then ψ is the conjunction of the following sentences:

- ($<$ is a linear order),
- ($(P_a)_{a\in\Sigma}$ form a partition),
- $\forall x_1 ... x_n \in P[\varphi_0(x_1...x_n) \wedge \bigwedge_{a\in\Gamma}(Q_a(x_1) \leftrightarrow \bigvee_{c\in f^{-1}(a)} P_c(x_1))]$, where $\varphi = \forall x_1...x_n \varphi_0(x_1...x_n)$ with φ_0 an open formula,
- ($(P_a)_{a\in\Sigma'}$ form a partition of $\neg P$),
- $\forall x_1...x_j[(\bigvee_{1\leq i\leq j} \neg P(x_i)) \to f(x_1...x_j) = x_1]$, for each j-ary function symbol f of $S(\varphi)$,

The sentence ψ is equivalent to a universal sentence, and it is local because in its models closure takes at most $n_\psi = n_\varphi$ steps. And $L(\psi) = f^{-1}(L(\varphi))$ holds by construction.

d) Closure under inverse morphism, intersection, complementation.
The proof uses the notion of rational cone of which we now recall the definition:

Definition 3.4 *A rational cone is a class of languages which is closed under morphism, inverse morphism, and intersection with a rational language. (Or, equivalently to these three properties, closed under rational transduction).*

Then recall the following:

Proposition 3.5 ([HU69], exercise) *A class of languages closed under morphism, inverse morphism, union and concatenation product, is a rational cone.*

Now we can state the next Proposition:

Proposition 3.6 *The class LOC is not a rational cone.*

Proof. It uses the following result:

Proposition 3.7 ([FZV80]) *The Antidyck language $Q_2'^{\star}$ is a generator of the rational cone of the recursively enumerable languages.*

But we have seen that the Antidyck language $Q_2'^{\star}$ is local, but there exist many recursively enumerable languages which are not local, for example $\{a^{2^n}/n \geq 1\}$, or the language ${}^cL$ of the above example 2.19.

Then from Propositions 3.5 and 3.6, we infer that LOC is not closed under inverse morphism. Now recall the following:

Theorem 3.8 (Nivat [Ber79]) *A class of languages which is closed under alphabetic morphism, inverse alphabetic morphism, and intersection with a rational language, is a rational cone.*

LOC is a class which is closed under alphabetic morphism, inverse alphabetic morphism but which is not a rational cone, therefore it is not closed under intersection with a rational language, and it is also neither closed under intersection, nor under complementation.

d) Closure under quotient.
Recall that the left quotient of L_1 by L_2 is $L_2\backslash L_1 = \{w/yw \in L_1$ for an $y \in L_2\}$.
And the right quotient of L_1 by L_2 is $L_1/L_2 = \{w/wy \in L_1$ for an $y \in L_2\}$.

Let then $R = a\{b^i a^i/i > 0\}^{\star}$ and $L = \{a^i b^{2i}/i > 0\}^{\star}$.
R and L are local and context free languages and it holds that:
$R\backslash L \cap b^+ = \{b^{2^n}/n > 0\}$ then $R\backslash L$ is not local. Because if $R\backslash L = L(\varphi)$ was local, the formula $\varphi \wedge \forall x P_b(x)$ would be local and would define $R\backslash L \cap b^+$, but $\{b^{2^n}/n > 0\}$ cannot be defined by a local sentence .
The proof is similar for the right quotient.

e) Closure under quotient by a finite language.
whenever the language is finite and contains a single word a of length 1, the language $\{a\}\backslash L(\varphi)$ is local, by projection of the models of φ on a predicate.
Furthermore we then use the formulas:
$L_1L_2\backslash L = L_2\backslash(L_1\backslash L)$ and $(L_1 \cup L_2)\backslash L = (L_1\backslash L) \cup (L_2\backslash L)$.

4 UNDECIDABLE PROBLEMS

Return to **Chapter 4** of [Gin66] about undecidable problems for context free languages . Showing that languages which appear there are local, we can prove the following undecidability results about local languages.

Theorem 4.1 *Let Σ be an alphabet containing at least two letters. It is undecidable to determine for arbitrary local languages $L(\varphi)$ and $L(\psi)$ over Σ whether :*

1. $L(\varphi) \cap L(\psi)$ *is empty.*

2. $L(\varphi) \cap L(\psi)$ *is infinite.*

3. $L(\varphi) \cap L(\psi)$ *is rational.*

4. $L(\varphi) \cap L(\psi)$ *is context free.*

5. $L(\varphi) \cap L(\psi)$ *is local: more precisely, there does not exist any algorithm which answers , either $L(\varphi) \cap L(\psi)$ is not local, or $L(\varphi) \cap L(\psi)$ is local, giving θ local and n_θ such that $L(\varphi) \cap L(\psi) = L(\theta)$. (closure taking at most n_θ steps in the models of θ).*

6. $L(\varphi) \subseteq L(\psi)$.

7. $L(\varphi) = L(\psi)$.

8. $L(\varphi) = \Sigma^\star$.

9. $L(\varphi)$ *is rational.*

10. ${}^cL(\varphi)$, *is empty . (${}^cL(\varphi) = \Sigma^\star - L(\varphi)$ being the complement of $L(\varphi)$ in $\Sigma^\star$).*

11. ${}^cL(\varphi)$ *is rational.*

12. ${}^cL(\varphi)$ *is context free.*

13. ${}^cL(\varphi)$ *is infinite.*

14. ${}^cL(\varphi)$ *is local . (with the same precision as for 5)).*

15. $L(\varphi)$ *is a linear context free language . (Σ has here at least three letters).*

16. *$L(\varphi)$ contains an infinite rational language.*

For arbitrary local sentence φ and rational language R:

17. *$L(\varphi) \supseteq R$.*

18. *$L(\varphi) = R$.*

Proof. We return to the **Post correspondance Theorem** :

Theorem 4.2 *Let Σ be an alphabet with at least two elements . Then it is undecidable to determine for arbitrary n-tuples $(w_1, .., w_n)$ and $(y_1, ...y_n)$ of nonempty words in $\Sigma^\star$ whether there exists a nonempty sequence of indices $i_1, ..., i_k$ such that $w_{i_1}....w_{i_k} = y_{i_1}...y_{i_k}$.*

Let $\Sigma = \{a, b, c\}$. For all n-tuples $x = (x_1, ...x_n)$ and $y = (y_1, ..y_n)$ of nonempty words of $\{a, b\}^\star$, let:
$L(x) = \{ba^{i_k}...ba^{i_1}cx_{i_1}...x_{i_k}/k \geq 1, 1 \leq i_j \leq n\}$ and $L(x, y) = L(x)cL(y)^R$, where for a language L, $L^R = \{u^R/u \in L\}$ and $u^R = u_n...u_2u_1$ whenever $u = u_1u_2...u_n$, u_i being a letter $\forall i$.

Let $\Gamma = \{c, a_1, ...a_n, c_1, ...c_n\}$ a new alphabet. And let h be the morphism $\Gamma^\star \to \Sigma^\star$ defined by : $c \to c, a_j \to x_j, c_j \to ba^j$. And let:
$L = \{c_{i_k}...c_{i_1}ca_{i_1}...a_{i_k}/k \geq 1, 1 \leq i_j \leq n\}$. L is a local language defined by the following sentence ψ of signature $S(\psi) = \{<, (P_{a_i})_{1\leq i\leq n}, P_c, (P_{c_i})_{1\leq i\leq n}, d, s\}$, where P_{a_i}, P_c, P_{c_i} are unary predicate symbols, d is a constant symbol and s is a unary function symbol.

ψ is the conjunction of :

- ($<$ is a linear order)
- $(P_{a_i}, P_c, P_{c_i}, 1 \leq i \leq n)$ form a partition,
- $\forall x(P(x) \leftrightarrow \bigvee_{1\leq i\leq n} P_{c_i}(x))$,
- $\forall x(Q(x) \leftrightarrow \bigvee_{1\leq i\leq n} P_{a_i}(x))$,
- $\forall xyz(P(x) \wedge Q(y) \wedge P_c(z) \to x < z < y)$,
- $\forall xy(P_c(x) \wedge P_c(y) \to x = y)$,
- $\forall x(s(s(x)) = x)$,
- $\forall x(P_{a_i}(x) \leftrightarrow P_{c_i}(s(x))), 1 \leq i \leq n$,

- $\forall x(P_c(x) \rightarrow s(x) = x)$,
- $P_c(d)$,
- $\forall xy(P(x) \wedge P(y) \wedge x < y \rightarrow s(y) < s(x))$.

Then the language L is local, and $L(x) = h(L)$ is also local because the image of a local language by a morphism is local. $L(x, y)$ is a local language: We easily see that A local implies A^R local from what we deduce that $L(y)^R$ is local and by concatenation product that $L(x).c.L(y)^R$ is local

Define now $L_s = \{w_1 c w_2 c w_2^R c w_1^R / w_1 \text{ and } w_2 \in \{a, b\}^\star\}$, $\Sigma = \{a, b, c\}$.

Let us prove that L_s is a local language: The language $\{a, b\}^\star c\{a, b\}^\star$ is rational then local and it is defined by the sentence $\tau = [\forall x(P_c(x) \rightarrow x = d)] \wedge [P_c(d)] \wedge [P_a, P_b, P_c$ form a partition $] \wedge [<$ is a linear order], where d is a constant symbol.

The language L_s is defined by ψ_s which is the conjunction of :

- ($<$ is a linear order),
- (P_a, P_b, P_c, form a partition),
- (P, Q, R, form a partition) ,
- $\forall xy(R(x) \wedge R(y) \rightarrow x = y \wedge P_c(x))$,
- $\forall xyz(P(x) \wedge Q(y) \wedge R(z) \rightarrow x < z < y)$,
- $P(d)$,
- $\forall x(s(s(x)) = x)$,
- $\forall x(P(x) \rightarrow Q(s(x)))$,
- $\forall x(R(x) \rightarrow s(x) = x)$,
- $\forall x \in P(P_c(x) \leftrightarrow x = d)$,
- $\forall xy(P(x) \wedge P(y) \wedge x < y \rightarrow s(y) < s(x))$
- $\forall x(P_a(x) \leftrightarrow P_a(s(x)))$,
- $\forall x(P_b(x) \leftrightarrow P_b(s(x)))$,

- $\forall x(P_c(x) \leftrightarrow P_c(s(x)))$.

With $S(\psi_s) = S(\tau) \cup \{P, Q, R, s\}$, where P, Q, R are unary predicate symbols and s is a unary function symbol. ψ_s is equivalent to a universal sentence and closure takes at most two steps in every model of it, hence ψ_s is local.

$L(x, y) \cap L_s$ consists of all words in the following form:
$ba^{i_k}...ba^{i_1}cx_{i_1}...x_{i_k}cy_{i_k}^R...y_{i_1}^Rca^{i_1}b...a^{i_k}b$ where:
$k \geq 1, 1 \leq i_j \leq n$ and $x_{i_1}...x_{i_k} = y_{i_1}...y_{i_k}$

Then $L(x, y) \cap L_s$ is empty if and only if there is not any solution to Post Correspondence Problem for x, y. In the other case $L(x, y) \cap L_s$ is infinite. From this fact we can deduce a), b), and c):

It is undecidable to determine , for arbitrary $L(x, y)$:

a) Whether $L(x, y) \cap L_s$ is empty.

b) Whether $L(x, y) \cap L_s$ is infinite.

c) Whether $\tau(L(x, y)) \cap \tau(L_s)$ is empty, where τ is the morphism $\{a, b, c\}^\star \rightarrow \{a, b\}^\star$ defined by : $a \rightarrow bab$, $b \rightarrow ba^2b$, $c \rightarrow ba^3b$.

c) results from the fact that $\tau(L(x, y)) \cap \tau(L_s)$ is empty iff $L(x, y) \cap L_s$ is empty.

Lemma 4.3 *$\tau(L(x, y)) \cap \tau(L_s)$ doesn't contain any infinite context free language.*

Proof. In [Gin66]

So we can deduce that $\tau(L(x, y)) \cap \tau(L_s)$ is context free iff it is rational iff it is empty iff it is not infinite.
$L(x, y)$ and L_s being local languages, and the image of a local language by a morphism being a local language, $\tau(L(x, y))$ and $\tau(L_s)$ are local languages over the alphabet $\{a, b\}$. ¿From this we deduce 1),2), 3), and 4) of our Theorem , because one cannot decide whether $\tau(L(x, y)) \cap \tau(L_s)$ is empty, infinite, rational or context free.

Let us show 5):
Suppose there exists an algorithm which, for any arbitrary given local languages $L(\varphi)$ and $L(\psi)$ over an alphabet $\Sigma = \{a, b\}$, answers: either $L(\varphi) \cap$

$L(\psi)$ is not local , either $L(\varphi) \cap L(\psi) = L(\theta)$ giving a local sentence θ and n_θ.

Then there are two cases:

1. **First case.** $L(\varphi) \cap L(\psi)$ is not local, and then $L(\varphi) \cap L(\psi)$ is not empty because the emptyset is a local language.

2. **Second case.** $L(\varphi) \cap L(\psi) = L(\theta)$, and then following [Res88] we could decide whether $L(\theta)$ is empty so whether $L(\varphi) \cap L(\psi)$ is empty. The proof of 1) shows there would be a contradiction, so 5) is proved.

Let us prove 6) –14).
First show that $\{a, b, c\}^\star - L(x, y)$ is a local language. It is the union of six languages $M_i, 1 \leq i \leq 6$. Let $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ and let:

$M_1 = \{a, b\}^\star \cup \{a, b\}^\star c\{a, b\}^\star \cup [\{a, b\}^\star c]^2\{a, b\}^\star \cup [\{a, b\}^\star c]^4\{a, b, c, \}^\star$

$\{a, b, c, \}^\star - M_1$ consists of all the words of $\{a, b, c, \}^\star$ with exactly three occurrences of c.

Let $M_2 = \{a, c\}\{a, b, c\}^\star \cup \{a, b, c\}^\star\{a, c\} \cup \{a, b, c\}^\star c^2\{a, b, c\}^\star$.

Then M_2 consists of all the words beginning by a or c or finishing by a or c or containing c^2 as a subword.

Let $M_3 = \{a, b\}^\star b^2\{a, b\}^\star c\{a, b, c\}^\star \cup \{a, b, c\}^\star c\{a, b\}^\star b^2\{a, b\}^\star \cup \{a, b\}^\star bc\{a, b, c\}^\star \cup \{a, b, c\}^\star cb\{a, b\}^\star$.

M_3 consists of all the words with either b^2 before the first c, either b^2 after the last c , either b immediately on the left of the first c , either immediately on the right of the last c.

Let $M_4 = \{a, b\}^\star a^{n+1}\{a, b\}^\star c\{a, b, c\}^\star \cup \{a, b, c\}^\star c\{a, b\}^\star a^{n+1}\{a, b\}^\star$.

M_4 consists of all the words with some $a^h, h \geq n + 1$, either before the first c or after the last c.

Each language $M_i, 1 \leq i \leq 4$, is rational, hence local.

Let $H = \{a, b, c\}^\star - \cup_{i=1}^4 M_i$.
H consists of all the words in the following form: $ba^{i_k}...ba^{i_1}cucvca^{j_1}b...a^{j_m}b$,

where u and v are nonempty words of $\{a,b\}^\star$, $k \geq 1, m \geq 1, 1 \leq i_r \leq n, 1 \leq j_s \leq n$.

$\{a,b,c\}^\star - H = \cup_{i=1}^{4} M_i$ is rational then it is local.

We are going to add two languages M_5 and M_6 to $\{a,b,c\}^\star - H$ in order that the resulting union be $\{a,b,c\}^\star - L(x,y)$.

Define for every nonempty word w of $\{a,b\}^\star$:
$D(w) = \{u \neq \lambda / u \in \{a,b\}^\star, |u| < |w|\}$ and
$J(w) = \{u \neq \lambda / u \in \{a,b\}^\star, u \neq w, |u| = |w|\}$
Next for each n-tuple $w = (w_1, ...w_n)$ of nonempty words w_i, we define:
$M(w) = c\{a,b\}^\star\{a,b\} \cup b\{a,b\}^\star c \cup \cup_{1 \leq i \leq n} \cup_{u \in D(w_i)} (ba^i cu \cup ba^i b\{a,b\}^\star cu) \cup \cup_{1 \leq i \leq n} \cup_{u \in J(w_i)} (ba^i c\{a,b\}^\star u \cup ba^i b\{a,b\}^\star c\{a,b\}^\star u)$.

$M(x)$ is a rational language (because for each i, $D(w_i)$ and $J(w_i)$ are finite sets), hence local and generated by a grammar $G = (V, \Sigma, P, \sigma)$. Let $G' = (V', \Sigma, P', \sigma')$, where σ' is not in V and $P' = P \cup \{\sigma' \rightarrow \sigma, \sigma' \rightarrow ba^i \sigma' x_i / 1 \leq i \leq n\}$.

It holds that $M_5 = L(G').c.\{a,b,c\}^\star$.
Remark that if h is the substitution $\{a,b,c\}^\star \rightarrow P(\{a,b,c\}^\star)$ defined by $a \rightarrow \{a\}$, $b \rightarrow \{b\}$, and $c \rightarrow M(x)$ then $h(L(x)) = L(G')$. But $L(x)$ and $M(x)$ being local, and local languages being closed under substitution, we can infer that $L(G')$ is local and then by concatenation product that M_5 is a local language.

Each word of $L(G')$ contains exactly one occurrence of c, $L(G') \cap L(x)$ is empty and $L(G')$ contains the set of words $\{ba^{i_k}...ba^{i_1}cw / w \neq x_{i_1}...x_{i_k}, w \in \{a,b\}^\star\}$.

$M_5 \cap H$ consists of all the words of H in the form $ba^{i_k}...ba^{i_1}cucvca^{j_1}b...a^{j_m}b$ with $u \in \{a,b\}^\star$ and $u \neq x_{i_1}...x_{i_k}$.

In a similar manner , let $G_1 = (V_1, \Sigma, P_1, \sigma_1)$ be a grammar generating $M(y)^R$. Let $G_2 = (V_1', \Sigma, P_1', \sigma_1')$, where σ_1' is a new symbol not in V_1 and $P_1' = P_1 \cup \{\sigma_1' \rightarrow \sigma, \sigma_1' \rightarrow y_i^R \sigma_1' a^i b / 1 \leq i \leq n\}$. Then $M_6 = \{a,b,c\}^\star.c.L(G_2)$.

And if h' is the substitution $\{a,b,c\}^\star \rightarrow P(\{a,b,c\}^\star)$ defined by $a \rightarrow \{a\}$, $b \rightarrow \{b\}$, and $c \rightarrow M(y)^R$, then $L(G_2) = h'(L(y)^R)$.

$L(y)^R$ and $M(y)^R$ are local languages therefore $L(G_2)$ is local by substitution and we deduce by concatenation product that M_6 is a local language.

$M_6 \cap H$ consists of all the words of H in the form $ba^{i_k}...ba^{i_1}cucvca^{j_1}b...a^{j_m}b$ with $v^R \neq y_{j_1}...y_{j_m}$.

Therefore $[\{a,b,c\}^\star - H] \cup M_5 \cup M_6 = \cup_{i=1}^{6} M_i = \{a,b,c\}^\star - L(x,y)$ is a local language.

Then considering the previous morphism τ, $\tau[\{a,b,c\}^\star - L(x,y)]$ is a local language, which is equal to $\tau[\{a,b,c\}^\star] - \tau[L(x,y)]$.
The language $\{a,b\}^\star - \tau[\{a,b,c\}^\star]$ is a rational language hence it is also a local one. But local languages are closed under union, then $\{a,b\}^\star - \tau[L(x,y)]$ is also a local language. And if Σ contains at least two elements, then $\Sigma^\star - [L(x,y)]$ is a local language.

Let us show now that $\{a,b\}^\star - \tau[L_s]$ is a local language:

$\{a,b,c\}^\star - L_s = M'_1 \cup M'_2 \cup M'_4$, Where:

M'_1 is the set words of $\{a,b,c\}^\star$ with less than three or more than three occurrences of c. M'_1 is rational then local.

M'_2 is the set of words $w_1cw_2cw_3cw_4$ where each $w_i \in \{a,b\}^\star$ and where $w_4 \neq w_1^R$. This language is the union of the following languages :

- $M_{2,1} = \{w_1cw_2cw_3cw_4 / w_i \in \{a,b\}^\star$ and $|w_1| < |w_4|\}$,
- $M_{2,2} = \{w_1cw_2cw_3cw_4 / w_i \in \{a,b\}^\star$ and $|w_4| < |w_1|\}$,
- $M_{2,3} = \{w_1cw_2cw_3cw_4 / w_i \in \{a,b\}^\star$ and $|w_1| = |w_4|$ and $w_4 \neq w_1^R\}$.

$M_{2,1}$ is defined by the following local sentence $\psi_{2,1}$, conjunction of:

- ($<$ is a linear order)
- (P_a, P_b, P_c, form a partition),
- (P, Q, R, form a partition) ,
- $\forall xyz(P(x) \wedge Q(y) \wedge R(z) \rightarrow x < z < y)$,
- $\forall x(P(x) \rightarrow P_a(x) \vee P_b(x))$,

- $\forall x(Q(x) \rightarrow P_a(x) \vee P_b(x))$,
- $\forall x(R(x) \rightarrow c_1 \leq x \leq c_3)$,
- $R(c_1) \wedge R(c_2) \wedge R(c_3) \wedge c_1 \neq c_2 \wedge c_1 \neq c_3 \wedge c_2 \neq c_3$,
- $\forall x(P_c(x) \leftrightarrow x = c_1 \vee x = c_2 \vee x = c_3)$,
- $\forall xy((P(x) \wedge P(y) \wedge x \neq y) \rightarrow i(x) \neq i(y))$,
- $\forall x(P(x) \rightarrow Q(i(x)))$,
- $\forall x(R(x) \vee Q(x) \rightarrow i(x) = x)$,
- $Q(d) \wedge \forall x(P(x) \rightarrow i(x) \neq d)$.

This sentence is equivalent to a universal sentence, its signature is $S(\psi_{2,1}) = \{<, P_a, P_b, P_c, P, Q, R, i, c_1, c_2, c_3, d\}$ where P, Q, R, are unary predicate symbols, i is a unary function symbol, and c_1, c_2, c_3, d , are constant symbols. And if $M \models \psi_{2,1}$ and $X \subseteq |M|$ then $cl(X, M) = cl^1(X, M)$.Therefore $\psi_{2,1}$ is local and the language $L(\psi_{2,1}) = M_{2,1}$ is local.

In a similar manner, $M_{2,2} = M_{2,1}^R$ is a local language.

$M_{2,3}$ is defined by the following local sentence $\psi_{2,3}$, conjunction of:

- ($<$ is a linear order),
- (P_a, P_b, P_c, form a partition),
- (P, Q, R, form a partition),
- $\forall xyz(P(x) \wedge Q(y) \wedge R(z) \rightarrow x < z < y)$,
- $\forall x(P(x) \rightarrow P_a(x) \vee P_b(x))$,
- $\forall x(Q(x) \rightarrow P_a(x) \vee P_b(x))$,
- $\forall x(R(x) \rightarrow c_1 \leq x \leq c_3)$,
- $R(c_1) \wedge R(c_2) \wedge R(c_3) \wedge c_1 \neq c_2 \wedge c_1 \neq c_3 \wedge c_2 \neq c_3$,
- $\forall x(P_c(x) \leftrightarrow x = c_1 \vee x = c_2 \vee x = c_3)$,
- $\forall x(i(i(x)) = x)$,
- $\forall x(P(x) \leftrightarrow Q(i(x)))$,

- $\forall x(R(x) \rightarrow i(x) = x)$,
- $\forall xy(P(x) \wedge P(y) \wedge x < y \rightarrow i(y) < i(x))$,
- $P_a(d) \wedge P_b(i(d))$.

the signature of $\psi_{2,3}$ is the same as $S(\psi_{2,1})$. $\psi_{2,3}$ is equivalent to a universal sentence and closure takes at most two steps in each of its models. So this sentence is local and the language $L(\psi_{2,3}) = M_{2,3}$ is local.

Then by union the language M'_2 is local.

Then $M'_4 = \{a,b\}^\star.c.M'_3.c.\{a,b\}^\star$, where M'_3 is the set of words w_2cw_3 , where w_2 and $w_3 \in \{a,b\}^\star$ and $w_3 \neq w_2^R$. By analogous methods as in the case of M'_2, We show that M'_3 is local , then by concatenation we can deduce that M'_4 is local.

Therefore, as the union of the local languages M'_1, M'_2 and M'_4, the language $\{a,b,c\}^\star - L_s$ is local. Then by an analogous reasoning as in the case of $\{a,b\}^\star - \tau[L(x,y)]$, we show that $\{a,b\}^\star - \tau[L_s]$ is a local language, Where τ is the above morphism.

Now , For Σ an alphabet containing at least two elements a and b, let: $M_1(x,y) = \Sigma^\star - (\tau[L(x,y)] \cap \tau[L_s]) = (\Sigma^\star - \tau[L(x,y)]) \cup (\Sigma^\star - \tau[L_s])$.

$M_1(x,y)$ is a local language as the union of two local languages, and $\Sigma^\star - M_1(x,y) = \tau[L(x,y)] \cap \tau[L_s]$

Then from our proof of 1), 2), 3), and 4), we can deduce 8), 10), 11), 12) and 13) because $\tau[L(x,y)] \cap \tau[L_s]$ is context free iff it is rational iff it is empty iff it is not infinite, and this is undecidable.

And $M_1(x,y)$ is a rational language iff $\Sigma^\star - M_1(x,y)$ is a rational language because the class of rational languages is closed under complementation, and this implies 9).

For 14) we reason as for 5).

8) implies 7) which implies 6).

8) implies also 17) and 18).

To prove 15) we follow [Gin66, exercise 16, p.128].
Let Σ containing at least three elements and let $c \in \Sigma$. For an arbitrary context free language $L \subseteq (\Sigma - \{c\})^\star$, LcL is a linear language iff L is rational. Then when $M_1(x, y) = \{a, b\}^\star - (\tau[L(x, y)] \cap \tau[L_s])$ it is undecidable to determine whether the local language $M_1(x, y)cM_1(x, y)$ is a linear language.

To prove 16) let us utilize **Lemma 4.3.4. of [Gin66]** :
For some n-tuples of nonempty words of $\{a, b\}^\star$: $y = (y_1, ..., y_n)$ and $z = (z_1, ..., z_n)$, consider the language $M(y, z) = \cup_{1 \leq r < \infty}[dL(y)]^r[dL(z)]^r$.
It is undecidable to determine whether the language $M(y, z)$, (over an alphabet containing at least four elements a, b, c, d) , contains an infinite rational language.
Then by Lemma 4.3.4. of [Gin66] , if τ' is the morphism $\{a, b, c, d\}^\star \rightarrow \{a, b\}^\star$ defined by $\tau'(a) = bab$, $\tau'(b) = ba^2b$, $\tau'(c) = ba^3b$, $\tau'(d) = ba^4b$, it is undecidable to determine whether the language $\tau'[M(y, z)]$ over $\{a, b\}$ contains an infinite rational language.
Then it suffices to show that $\tau'[M(y, z)]$ is a local language and, because the image of a local language by a morphism is local, that $M(y, z)$ is local.
The language $L = \{e^n f^n / 1 \leq n\}$ is local . And if k is the substitution $\{e, f\}^\star \rightarrow P(\{a, b, c, d\}^\star)$ defined by $e \rightarrow dL(y)$ and $f \rightarrow dL(z)$, then $k(L) = M(y, z)$. Therefore $M(y, z)$ is local because L is local and local languages are closed under substitution.

5 Definitions and review of transfinite length word languages

5.1 words of infinite and transfinite length

We shall assume the reader to be familiar with the elementary theory of countable ordinals, which may be found in [Sie65].
Let Σ be a finite alphabet, and α be an ordinal. A word of length α (an α-word) over the alphabet Σ is an α-sequence (or sequence of length α) of letters in Σ. In a similar manner as in the case of finite words, we' ll identify a word σ of length α over Σ with the structure $(\alpha, <^\sigma, (P_a^\sigma)_{a\in\Sigma})$ which signature is $\Lambda_\Sigma = \{<, (P_a)_{a\in\Sigma}\}$ where the unary predicate P_a is interpreted by $P_a^\sigma = \{i < \alpha$ such that the $i + 1^{th}$ letter of σ is an $a\}$.
Let then α be an ordinal $\geq \omega$. We' ll denote Σ^α the set of α-words over the alphabet Σ. And we define an α-language over Σ as a subset of Σ^α. We can

now set the following definition:

Definition 5.1 *Let L be an α-language over the alphabet Σ. L is a locally finite α-language iff there exists a local sentence φ in a signature $\Lambda \supseteq \Lambda_\Sigma$ such that: $(\sigma \in L) \longleftrightarrow (\exists M, M \models \varphi, M$ of order type α and $M|\Lambda_\Sigma = \sigma)$. (where $M|\Lambda_\Sigma$ is the reduction of M to the signature Λ_Σ).*

Notation. Let us denote $L^\Sigma_\alpha(\varphi)$ (or $L_\alpha(\varphi)$ if there is no ambiguity about the alphabet Σ) the α-language defined by the local sentence φ. And we' ll note $L_{\geq\alpha}(\varphi) = \cup_{\beta\geq\alpha} L_\beta(\varphi)$ and $L_{>\alpha}(\varphi) = \cup_{\beta>\alpha} L_\beta(\varphi)$.

5.2 Automata over words of length ω

Recall first the notion of Büchi automaton reading words of length ω, [Eil74], [Tho90], [PP98].
Intuitively a Büchi automaton is a finite automaton which reads infinite words, the acceptation condition for an ω-word being that during its reading (or during one of its readings, in the non deterministic case), one of the "final" states appears infinitely often.

Definition 5.2 *A Büchi automaton over the alphabet Σ is of the form $Aut = (Q, q_0, \Delta, F)$ where Q is a finite set of states, q_0 is the initial state $\in Q$, Δ is the transition relation ($\Delta \subseteq Q \times \Sigma \times Q$) and F is a subset of Q called the set of final states.*
A run of Aut reading an ω-word $\sigma = \sigma(0)\sigma(1)...\sigma(n)...$ of Σ^ω is an ω-sequence $\alpha = \alpha(0)\alpha(1)...\alpha(n)...$ such that $\alpha(0) = q_0$ and $(\alpha(i), \sigma(i), \alpha(i+1)) \in \Delta$ for $i \geq 0$.
The run is called successful if $Inf(\alpha) \cap F \neq \emptyset$,where $Inf(\alpha)$ is the set of elements of Q which appear infinitely often in the ω-sequence α. The automaton Aut accepts the ω-word σ if there exists a successful run of Aut over σ.We denote $L_\omega(Aut) = \{\sigma \in \Sigma^\omega \setminus Aut$ accepts $\sigma\}$ the ω-language recognized by Aut. If $L = L_\omega(Aut)$ for a Büchi automaton Aut, L is called a Büchi ω-language .

Recall now some of the essential results about Büchi ω-languages:

Theorem 5.3 *The class of Büchi ω-languages is the omega-Kleene closure of the set of rational languages , i.e. the class of languages in the form $\cup_{i=1}^n U_i.V_i^\omega$,where n is a non negative integer, and for each i∈[1,n], U_i and V_i are rational languages.*

Theorem 5.4 ([Tho90]) *The emptiness problem for Büchi ω-languages is decidable. (this means that: "For an arbitrary Büchi automaton Aut, is $L_\omega(Aut)$ empty? " is a decidable problem.)*

Remark 5.5 *Muller defined another class of automata reading words of length ω. the difference between Büchi and Müller automata is the notion of acceptation of an ω-word . A Muller automata is in the form $Aut = (Q, q_0, \Delta, F)$ where Q, q_0, Δ are defined in a similar manner as in the Büchi case and where $F \subseteq P(Q)$, (F is a subset of the power set of Q). A run of the automata is defined as above but a run α of Aut over σ is successful if $Inf(\alpha) \in F$.*

Theorem 5.6 *(Mac Naughton). An ω-language over the alphabet Σ is Muller recognizable iff it is Büchi recognizable.*

The advantage of Muller automata is that the deterministic version has the same expressive power as the non deterministic version, while this is false for Büchi automata.

5.3 Büchi ω-languages are local ω-languages

We shall show the following result:

Theorem 5.7 *Every Büchi ω-language is a local ω-language and, if B is a Büchi ω-language, there exists a local sentence φ such that $L_\omega(\varphi) = B$ and $L_{>\omega}(\varphi) = \emptyset$.*

Proof. We shall use the characterization of Büchi ω-languages by ω-regular expresssions. Let B be a Büchi ω-language over the alphabet Σ. Then there exist some rational languages (U_i) and (V_i) over Σ, where $1 \leq i \leq n$, such that $B = \cup_{i=1}^{n} U_i.V_i^\omega$. Recall that for $V \subseteq \Sigma^\star$, we denote $V^\omega = \{\alpha \in \Sigma^\omega \setminus \alpha = \alpha_1\alpha_2\alpha_3...\alpha_n.....$ with $\alpha_i \in V, \forall i \geq 1\}$.

Recall The following Lemma [Fin89]: If U is a rational language over Σ, Then there exists a local sentence φ such that $U = L^\Sigma(\varphi)$ and $L_\alpha^\Sigma(\varphi) = \emptyset$ for α ordinal $\geq \omega$.
This is proved by induction on the complexity of a regular expression defining a rational language.

Let then some local sentences φ_i and ψ_i, $1 \leq i \leq n$, such that
$\forall i \in [1, n], L(\varphi_i) = U_i, L(\psi_i) = V_i$ and $L_{\geq\omega}(\varphi_i) = \emptyset$ and $L_{\geq\omega}(\psi_i) = \emptyset$.

Recall also that for two local sentences φ_1 and φ_2 defining local languages $L(\varphi_1)$ and $L(\varphi_2)$, we can easily define the local sentences $\varphi_1.\varphi_2$ and $\varphi_1 \cup \varphi_2$ such that $L(\varphi_1.\varphi_2) = L(\varphi_1).L(\varphi_2)$ and $L(\varphi_1 \cup \varphi_2) = L(\varphi_1) \cup L(\varphi_2)$.

Remark that for words of transfinite length, we get:

$$L_\alpha(\varphi_1.\varphi_2) = \cup_{(\beta_1+\beta_2=\alpha)} L_{\beta_1}(\varphi_1).L_{\beta_2}(\varphi_2)$$

$$L_\alpha(\varphi_1 \cup \varphi_2) = L_\alpha(\varphi_1) \cup L_\alpha(\varphi_2).$$

Recall also (see proof of Theorem 3.1) that for any local sentence φ,

$\varphi^\star$ is defined in the signature $\mathrm{S}(\varphi^\star)$ which is the signature of φ to which is added a unary function symbol I and in which every constant symbol e is replaced by a unary function symbol.

$\varphi^\star$ is the sentence defined by the conjunction of:

- ($<$ is a linear order),
- $\forall yz[I(y) \leq y$ and $(y \leq z \rightarrow I(y) \leq I(z))$ and $(I(y) \leq z \leq y \rightarrow I(z) = I(y))]$,
- $\forall xy[I(x) = I(y) \rightarrow e(x) = e(y)]$, for each constant e of the signature $\mathrm{S}(\varphi)$ of φ,
- $\forall x_1...x_n[(\bigvee_{i,j\leq n} I(x_i) \neq I(x_j)) \rightarrow f(x_1...x_n) = min(x_1...x_n)]$, for each n-ary function f of $\mathrm{S}(\varphi)$,
- $\forall x_1...x_n[(\bigwedge_{i,j\leq n} I(x_i) = I(x_j)) \rightarrow I(f(x_1...x_n)) = I(x_1)]$, for each n-ary function f of $\mathrm{S}(\varphi)$,
- $\forall x \varphi^x$, where φ^x is the local sentence φ in which every constant e is replaced by the term $e(x)$ and each quantifier is relativized to the set $\{y \setminus I(y) = I(x)\}$.

Remark 5.8 *The models of $\varphi^\star$ are essentially direct sums of models of φ , and $L^\Sigma(\varphi^\star) = (L^\Sigma(\varphi))^\star$.*

Remark 5.9 *These three operations over local sentences : $\varphi, \psi \rightarrow \varphi.\psi$ then $\varphi, \psi \rightarrow \varphi \cup \psi$ and $\varphi \rightarrow \varphi^\star$, already permit us to construct a sentence defining the Büchi language B.*

We have $B = \cup_{i=1}^{n} U_i.V_i^{\omega}$ and $U_i = L(\varphi_i), V_i = L(\psi_i)$ for $1 \leq i \leq n$. Let then the local sentence $\cup_{i=1}^{n} \varphi_i.(\psi_i^{\star}) = \varphi$. We easily verify that $L_{\omega}(\varphi) = \cup_{i=1}^{n} L(\varphi_i).L_{\omega}(\psi_i^{\star})$, i.e. $L_{\omega}(\varphi) = \cup_{i=1}^{n} L(\varphi_i).L(\psi_i)^{\omega} = \cup_{i=1}^{n} U_i.V_i^{\omega} = B$.

Remark 5.10 *The hypothesis $L_{\geq\omega}(\varphi_i) = \emptyset$ and $L_{\geq\omega}(\psi_i) = \emptyset$ was necessary to state that $L_{\omega}(\psi_i^{\star}) = L(\psi_i)^{\omega}$ and that $L_{\omega}(\varphi) = \cup_{i=1}^{n} L(\varphi_i).L_{\omega}(\psi_i^{\star})$.*

Remark 5.11 *Every Büchi ω-language is then a local ω-language. For the continuation, we'll show that we can obtain φ such that $L_{\omega}(\varphi) = B$ and $L_{>\omega}(\varphi) = \emptyset$. for that, let us define a new operation over local sentences.*

Definition 5.12 *Let φ and ψ local sentences such that $S(\varphi^{\star}) \cap S(\psi) = \{<\}$. We define the sentence $\varphi^{(\star\psi)}$ in the signature $S(\varphi^{\star}) \cup S(\psi) \cup \{P\}$, where P is a new unary predicate symbol . $\varphi^{(\star\psi)}$ is the conjunction of :*

- $\varphi^{\star}$,
- $\forall x[P(x) \leftrightarrow I(x) = x]$,
- $\forall x_1...x_n[(\bigwedge_{i=1}^{n} P(x_i)) \rightarrow \psi_1(x_1...x_n)]$, *where* $\psi = \forall x_1...x_n \psi_1(x_1...x_n)$ *and* ψ_1 *is an open formula,*
- $\forall x_1...x_k[(\bigwedge_{i=1}^{k} P(x_i)) \rightarrow P(t(x_1...x_k))]$, *for each k-ary function t of S(ψ),*
- $\forall x_1...x_k[Q(x_1...x_k) \rightarrow P(x_1) \wedge ... \wedge P(x_k)]$, *for each k-ary predicate symbol Q of S(ψ)*
- $P(a)$, *for each constant a of S(ψ),*
- $\forall x_1...x_n[(\bigvee_{i=1}^{n} \neg P(x_i)) \rightarrow t(x_1...x_n) = min(x_1...x_n)]$, *for each n-ary function t in S(ψ).*

Remark 5.13 *The models of $\varphi^{(\star\psi)}$ essentially are direct sums of models of φ , these models being ordered by the order type of a model of ψ.*

Recall now the following result:

Lemma 5.14 ([FR96]) *There exists a local sentence φ_{ω}, with $<\in S(\varphi_{\omega})$, which has a model of order type ω (for $<$) and no model of order type an ordinal $\alpha > \omega$.*

We can then modify the expression of φ we are looking for: Let $\varphi = \cup_{i=1}^{n} \varphi_i.\psi_i^{(\star\varphi_\omega)}$. We have then $L_\omega(\varphi) = \cup_{i=1}^{n} U_i.V_i^\omega = B$, and this time φ has not any model of order type $> \omega$.

Remark 5.15 *The same proof shows that every ω-language in the form $\cup_{i=1}^{n} U_i.V_i^\omega$ where U_i and V_i are local languages , is a local ω-language. then the ω-Kleene closure of the set of local languages is included in the set of local ω-languages. We have shown that local languages extend far beyond rational languages , that many context free languages are local, and many non context free also. So the above proof implies that local ω-languages extend far beyond Büchi ω-languages [Fin89] [Fin99].*

6 Automata over words of transfinite length

So as an automata to be able to read words of length $\geq \omega$, we must add to the automaton a transition relation for limit steps: after the reading of a word which length is a limit ordinal, the state of the automaton will depend on the set of states which cofinally appeared during the run of the automaton [BS73], [Hem92], [Bed96].

6.1 First definitions

Definition 6.1 *A generalized Büchi automaton is a sextuple $(\Sigma, Q, q_0, \Delta, \gamma, F)$ where: Σ is a finite alphabet , Q is a finite set of states, q_0 is a state in Q called initial state, $\Delta \subset Q \times \Sigma \times Q$, is the transition relation, $\gamma \subset P(Q) \times Q$.*

Σ, Q, q_0, Δ and F keep the same signification as before, the signification of γ is given by the :

Definition 6.2 *A run of the generalized Büchi automaton $Aut = (\Sigma, Q, q_0, \Delta, \gamma, F)$ reading the word σ of length α, is an (α+1)-sequence of states x defined by : $x(0) = q_0$ and for $i < \alpha$, $(x(i), \sigma(i), x(i+1)) \in \Delta$ and for i a limit ordinal, $(Inf(x, i), x(i)) \in \gamma$, where $Inf(x, i) = \{q \in Q \setminus \forall\mu < i, \exists\nu < i$ such that $\mu < \nu$ and $x(\nu) = q\}$.*
$Inf(x, i)$ is the set of states which cofinally appear during the reading of the i first letters of σ.
A run x of the automaton Aut over the word σ of length α is called successful if $x(\alpha) \in F$. A word σ of length α is accepted by Aut if there exists a successful run of Aut over σ. We denote $L_\alpha(Aut)$ the set of words of length α which are accepted by Aut. An α-language L is a (generalized) Büchi α-language if there exists a generalized Büchi automaton A such that $L = L_\alpha(A)$.

Remark 6.3 *When we consider only finite words, the language accepted by a generalized Büchi automaton is a rational language. And the notion of ω-language accepted by generalized Büchi automaton corresponds to that of ω-language accepted by Muller automaton and then also by Büchi automaton.*

6.2 ω^2-languages accepted by generalized Büchi automata are local ω^2-languages

Let $Aut = (\Sigma, Q, q_0, \Delta, \gamma, F)$ be a generalized Büchi automaton, and let $L_{\omega^2}(Aut)$ be the ω^2-language recognized by Aut.
We shall decompose the reading of an ω^2-word by Aut into blocks of length ω.
Let σ be an ω^2-word. Aut reads the word σ: after the reading of the first ω letters, Aut is in state $x(\omega)$, after the reading of $\omega.2$ letters, Aut is in state $x(\omega.2)$ and so on...
Define now the following ω-languages :
For $q_i \in Q, q_j \in Q$ and $E \subseteq Q$, $L(q_i, q_j, E)$ is the ω-language of words u such that there exists a reading of u by Aut such that Aut reads the word u, beginning in state q_i, it is in state q_j after the reading of u and the set of states in which the automaton has been (q_i and q_j comprised) is the set E.
We easily see that these ω-languages are recognized by Muller automata therefore also by Büchi automata.

Consider now a new alphabet: $\Gamma = \{(q_i, q_j, E) \setminus q_i \in Q, q_j \in Q, E \subseteq Q\} = Q \times Q \times P(Q)$ and consider the ω-language over Γ of the ω-words such that:
The first letter is in the form (q_0, q, E) and each letter (q_i, q_j, E) is followed by a letter (q_j, q, G) with $q \in Q, G \subseteq Q$. And such that $X = \{q \in Q \setminus$ there exists a letter (q_i, q_j, G) which appears infinitely often with $q \in G\}$ satisfies $(X, q_f) \in \gamma$ for a $q_f \in F$.
This ω-language over Γ is a Büchi ω-language. Denote it by L^{Γ}_{Aut}. Remark that if we substitute in L^{Γ}_{Aut} the ω-language $L(q_i, q_j, E)$ for each letter (q_i, q_j, E) , we obtain the ω^2-language recognized by Aut, i.e. $L_{\omega^2}(Aut)$.
From the preceding results, there exists a local sentence φ^{Γ} such that $L^{\Gamma}_{\omega}(\varphi^{\Gamma}) = L^{\Gamma}_{Aut}$ and $L_{>\omega}(\varphi^{\Gamma}) = \emptyset$. In the same way there exist local sentences $\varphi_{(q_i,q_j,E)}$ such that $L^{\Sigma}_{\omega}(\varphi_{(q_i,q_j,E)}) = L(q_i, q_j, E)$ and $L_{>\omega}(\varphi_{(q_i,q_j,E)}) = \emptyset$.
We then use the substitution method established in the above section :" closure properties of local languages ".
We'll now assume that the signatures of the sentences $\varphi_{(q_i,q_j,E)}$ satisfy:
$S(\varphi_{(q_i,q_j,E)}) \cap S(\varphi_{(q'_i,q'_j,E')}) = \{<, (P_a)_{a\in\Sigma}\}$ for $(q_i, q_j, E) \neq (q'_i, q'_j, E')$.

Let Q_a be the unary predicate in $S(\varphi^\Gamma)$, which indicates the place of the letter $a \in \Gamma$. We also assume, possibly differently naming the function, constant and other predicate symbols of $S(\varphi^\Gamma)$ that $S(\varphi_{(q_i,q_j,E)}) \cap S(\varphi^\Gamma) = \{<\}$ for every $(q_i, q_j, E) \in Q \times Q \times P(Q)$.

Let us now construct a local sentence defining the language $L_{\omega^2}(Aut)$:
ψ is the conjunction of the following sentences (the meaning of which is explained below):

- ($<$ is a linear order),
- $\forall xy[(I(y) \leq y) \wedge (y \leq x \rightarrow I(y) \leq I(x)) \wedge (I(y) \leq x \leq y \rightarrow I(x) = I(y))]$,
- $\forall x[I(x) = x \leftrightarrow P(x)]$,
- $P(c)$, for each constant c of $S(\varphi^\Gamma)$,
- $\forall x_1...x_k[R(x_1...x_k) \rightarrow P(x_1) \wedge .... \wedge P(x_k)]$, for each predicate $R(x_1...x_k)$ of $S(\varphi^\Gamma)$,
- $\forall x_1...x_j[(P(x_1) \wedge ... \wedge P(x_j)) \rightarrow P(f(x_1...x_j))]$, for each j-ary function symbol f of $S(\varphi^\Gamma)$,
- $\forall x_1...x_j[(\bigvee_{1\leq i\leq j} \neg P(x_i)) \rightarrow f(x_1...x_j) = min(x_1...x_j)]$, for each j-ary function symbol f of $S(\varphi^\Gamma)$,
- $\forall x_1...x_m[(P(x_1) \wedge ... \wedge P(x_m)) \rightarrow \varphi_0^\Gamma(x_1...x_m)]$, where $\varphi^\Gamma = \forall x_1...x_m \varphi_0^\Gamma(x_1...x_m)$ with φ_0^Γ an open formula,
- $\forall x_1...x_j[\bigvee_{i,k\leq j}(I(x_i) \neq I(x_k)) \rightarrow f(x_1...x_j) = min(x_1...x_j)]$, for every function f of $S(\varphi_{(q_i,q_j,E)})$ for $(q_i, q_j, E) \in Q \times Q \times P(Q)$,
- $\forall xy_1...y_j[(\bigwedge_{1\leq l\leq j} I(y_l) = I(x)) \rightarrow I(f(y_1...y_j)) = I(x)]$, for each j-ary function symbol f of $S(\varphi_{(q_i,q_j,E)})$ for $(q_i, q_j, E) \in \Gamma$,
- Finally, for each letter (q_i, q_j, E) and the associated sentence $\varphi_a = \varphi_{(q_i,q_j,E)}$:
- $\forall xy_1...y_p[(\bigwedge_{1\leq l\leq p} I(y_l) = I(x) \wedge Q_a(I(x))) \rightarrow \varphi_a^0(y_1...y_p) \wedge \{(e_b(y_1) = I(x) \wedge f_b(y_1...y_p) = y_1 \wedge \neg R_b(y_1...y_p) \wedge \bigwedge_{e_a \in S(\varphi_a)} e_a(y_1) = e_a(x)$; where $b \neq a$, and e_b, f_b, R_b run over the constants, functions, and predicates of $S(\varphi_b)\}]$,

Above , 1) to each constant e_a of $S(\varphi_a)$ is associated a new unary function $e_a(y)$ and 2) whenever $\varphi_a = \forall y_1...y_p \psi_a(y_1...y_p)$ with ψ_a open, φ_a^0 is ψ_a in which every constant e_a has been replaced by the function e_a.
At last we have the following conjunctions:

- $\forall xy[(P(x) \wedge P(y) \wedge x < y) \rightarrow (x < h(xy) < y \wedge \neg P(h(xy)))]$,
- $\forall xyz \in P[x < y < z \rightarrow x < h(xy) < h(xz) < y]$,
- $\forall xy[(\neg P(x) \vee \neg P(y)) \rightarrow h(xy) = x]$,
- $\forall xy[(P(x) \wedge P(y) \wedge \neg(x < y)) \rightarrow h(xy) = x]$.

Above h is a new binary function, P a unary predicate, I a unary function added to $S(\psi)$.

Construction of ψ :
Using the function I which marks the first letters of the subwords, we divide a word into subwords. In every model M of ψ, the set of the "first letters of subwords" , P^M, grows richer in a model of φ^Γ (therefore will constitute a word of L^Γ_{Aut}).
Then we "substitute" : for each letter (q_i, q_j, E) in P^M, we substitute a word of the associated language $L(q_i, q_j, E)$, using for that the formula $\varphi_{(q_i,q_j,E)}$.
The last four conjunctions imply that in a model M of ψ of order type ω^2, (P^M being of order type ω because the $\varphi_{(q_i,q_j,E)}$ have not any model of order type $> \omega$), every subword is of order type ω, and therefore is a word of the associated Büchi language $L(q_i, q_j, E)$. Then it really holds that $L_{\omega^2}(\psi) = L_{\omega^2}(Aut)$.

Remark 6.4 *This sentence is local because it is equivalent to a universal one, and closure takes a finite number of steps in a model of ψ: one takes the closure by I, then by the functions of $S(\varphi^\Gamma)$, then by h, afterwards by the functions of the $S(\varphi_{(q_i,q_j,E)})$.*

Remark 6.5 *By construction, the sentence ψ has not any well ordered model of order type $> \omega^2$, because φ^Γ and the $\varphi_{(q_i,q_j,E)}$ have no well ordered model of order type $> \omega$.*

6.3 ω^n-languages recognized by generalized Büchi automata are local ω^n-languages

We shall prove the following result:

Theorem 6.6 *Whenever n is an integer ≥ 1, every ω^n-language which is recognized by a generalized Büchi automaton is defined by a local sentence which has not any well ordered model of order type $> \omega^n$.*

Proof.
We reason by induction over n. The cases $n = 1$ and $n = 2$ were solved in preceding paragraphs.
Suppose the result be established for the integer $n - 1$.
Let then a generalized Büchi automata $Aut = (\Sigma, Q, q_0, \Delta, \gamma, F)$ reading words of length ω^n with $n \geq 2$.
We now use a method which is analogous to that one used in the preceding case:
We can divide a word of length ω^n into ω subwords of length ω^{n-1}.
Let as above the new alphabet $\Gamma = Q \times Q \times P(Q)$ and the ω-language over Γ: $L^{\Gamma}_{Aut} = L_{\omega}(\varphi^{\Gamma})$.
And let $L^{n-1}(q_i, q_j, E)$, the ω^{n-1}-language of the ω^{n-1}-words u such that if Aut reads the word u beginning the reading in state q_i , it finishes the reading in state q_j and the set of states entered by Aut during the run is E, (q_i and q_j comprised in E).
We easily see that the languages $L^{n-1}(q_i, q_j, E)$ are ω^{n-1}-languages recognized by generalized Büchi automata.
Then remark that if one substitute in the language L^{Γ}_{Aut} the
language $L^{n-1}(q_i, q_j, E)$ for the letter (q_i, q_j, E), we obtain the ω^n-language $L_{\omega^n}(Aut)$.
By induction assumption, each language $L^{n-1}(q_i, q_j, E)$ is defined by a local sentence $\varphi^{n-1}_{(q_i,q_j,E)}$: we have $L^{n-1}(q_i, q_j, E) = L^{\Sigma}_{\omega^{n-1}}(\varphi^{n-1}_{(q_i,q_j,E)})$.
We then use the substitution method as above: we obtain the sentence ψ of preceding paragraph, where we replaced the sentences $\varphi_{(q_i,q_j,E)}$ by the sentences $\varphi^{n-1}_{(q_i,q_j,E)}$ and the four last conjunctions by the conjunction of the following sentences (where h is a n-ary function symbol).

- $\forall x_1...x_n[(\bigvee_{1\leq i\leq n} \neg P(x_i)) \rightarrow h(x_1...x_n) = x_1]$,
- $\forall x_1...x_n[(\bigvee_{2\leq i\leq n} \neg(x_1 < x_i)) \rightarrow h(x_1...x_n) = x_1]$.
- $\forall x_1...x_n[(\bigwedge_{1\leq i\leq n} P(x_i) \wedge \bigwedge_{2\leq i\leq n}(x_1 < x_i)) \rightarrow \neg P(h(x_1...x_n))]$,

- $\forall x_1...x_n[(\bigwedge_{1\leq i\leq n} P(x_i) \wedge \bigwedge_{2\leq i\leq n}(x_1 < x_i) \wedge P(y) \wedge x_1 < y) \rightarrow x_1 < h(x_1x_2...x_n) < y]$,
- $\forall x_1...x_n\forall y_1...y_n[(\bigwedge_{1\leq i\leq n} P(x_i)\wedge\bigwedge_{1\leq i\leq n} P(y_i)\wedge\bigwedge_{2\leq i\leq n} x_1 < x_i\wedge\bigwedge_{2\leq i\leq n} x_1 < y_i \wedge \bigwedge_{1\leq i\leq j} x_i = y_i \wedge x_{j+1} < y_{j+1}) \rightarrow h(x_1...x_n) < h(y_1...y_n)]$, for each integer j such that $1 \leq j < n$.

The principle is the same as before: if M is a model of order type ω^n of ψ, P^M is of order type ω because every sentence $\varphi^{n-1}_{(q_i,q_j,E)}$ has no well ordered model of order type $> \omega^{n-1}$, and then the function h ensures that every subword is of length ω^{n-1}.
$M|\Lambda_\Sigma$ is then a word of $L_{\omega^n}(Aut)$ and the converse is true: every word of $L_{\omega^n}(Aut)$ grows richer in a model of ψ. Then it really holds that: $L^\Sigma_{\omega^n}(\psi) = L_{\omega^n}(Aut)$.
More, by construction, ψ has no well ordered model of order type $> \omega^n$. This achieves the proof by induction.

Remark 6.7 *The construction of the sentence ψ may be done in an effective manner from the automaton Aut.*

6.4 α-languages ($\omega \leq \alpha < \omega^\omega$) recognized by generalized Büchi automata are local α-languages

We shall prove the following

Theorem 6.8 *Let α be an ordinal such that $\omega \leq \alpha < \omega^\omega$. Then every α-language recognized by a generalized Büchi automaton is a local α-language.*

Proof. Let α be an ordinal such that $\omega \leq \alpha < \omega^\omega$. The ordinal α admits a decomposition into the Cantor normal form [**?**]:
$\alpha = \omega^{p_k}.n_k + \omega^{p_{k-1}}.n_{k-1} + ... + \omega^{p_1}.n_1 + n_0$
where p_i, n_i, k are integers such that $p_k > p_{k-1} > .... > p_1 \geq 1$ and $n_0 \geq 0, n_i \geq 1$ for $1 \leq i \leq k$, and $k \geq 1$.
let $Aut = (\Sigma, Q, q_0, \Delta, \gamma, F)$ be a generalized Büchi automaton reading α-words over Σ. We reason as above, dividing this time an α-word into n_k subwords of length ω^{p_k}, then n_{k-1} subwords of length $\omega^{p_{k-1}}$ and so on ... up to n_1 subwords of length ω^{p_1} and a finite subword of length n_0.
Assume $n_0 > 0$, the case $n_0 = 0$ being treated in a similar manner.
Let then $L^n(q_i, q_j)$ be the ω^n-language of the ω^n-words σ over Σ such that there exists a run of the automaton Aut reading the word σ where Aut begins

the reading of σ in state q_i and finishes it in state q_j (where $q_i \in Q, q_j \in Q$). And let R_q the rational language of finite words σ such that if Aut begins to read σ in state q, it finishes the reading in a final state of F.
Let then the (finite) language L of finite words, of length $n_k + n_{k-1} + ...n_1 + 1$ over the alphabet $A = \{L^{p_l}(q_i, q_j) \setminus q_i \in Q, q_j \in Q, 1 \leq l \leq k\} \cup \{R_q \setminus q \in Q\}$. L is constituted of the words which begin with a letter in the form $L^{p_k}(q_0, q_i)$, such that the n_k first letters are among the $L^{p_k}(q_i, q_l)$ and such that a letter $L^{p_k}(q_i, q_l)$ is followed by a letter $L^{p_k}(q_l, q_m)$. The following n_{k-1} letters are among the letters $L^{p_{k-1}}(q_i, q_l)$ then the n_{k-2} following letters are among the $L^{p_{k-2}}(q_i, q_l)$, until the n_1 letters among the $L^{p_1}(q_i, q_j)$ followed by a letter among the R_q, $q \in Q$. And always a letter $L^{p_m}(q_i, q_l)$ is followed by a letter $L^{p_m}(q_l, q_j)$ or by a letter $L^{p_{m-1}}(q_l, q_j)$, and the latest letter of the word being R_q if the last but one letter is in the form $L^{p_1}(q_i, q)$.
By construction, if one substitute in the language L for each letter $L^{p_m}(q_i, q_j)$ or R_q the associated language (over Σ), we obtain the α-language $L_\alpha(Aut)$. The language L is finite then it is local and it is defined by a local sentence φ which has no model of cardinal $> n_k + n_{k-1} + ...n_1 + 1$.
Each language $L^{p_j}(q_i, q_m)$ is recognized by a generalized Büchi automaton therefore (from preceding paragraph) it is defined by a local sentence $\varphi_{p_j}(q_i, q_m)$ such that $L_{>\omega^{p_j}}(\varphi_{p_j}(q_i, q_m)) = \emptyset$.
For $q \in Q$, the language R_q is rational then it is defined by a local sentence ψ_q such that $L_{\geq\omega}(\psi_q) = \emptyset$.
Again with the " substitution method ", we obtain a local sentence ψ_α such that $L_\alpha^\Sigma(\psi_\alpha) = L_\alpha(Aut)$.

6.5 Conclusion

We have proved in [FR96] the following:

Theorem 6.9 *let φ be a local sentence with a symbol $<$ in $S(\varphi)$, and α be an ordinal such that $\omega \leq \alpha < \omega^\omega$, has φ a well ordered model (for $<$) of order type α? is a decidable problem.*

The proof relies upon the existence of indiscernables in a model which reduces the existence of a model of order type α to the existence of a finite model of another local sentence which is effectively obtained from φ.
The preceding method, which associates a local sentence to a Büchi α-language then allows to decide the emptiness problem for a Büchi α-language. The α-languages which are recognized by finite automata were first studied by Büchi in order to obtain a decision algorithm for the monadic theory of

$(\alpha, <)$.
The preceding result (Theorem 6.9) allows to obtain a decision algorithm for the sentences in the form $\exists R_1....\exists R_k\varphi$, where φ is local in the signature $S(\varphi) = \{<, R_1, ...R_k\}$, where $R_1, ...R_k$ are relation or n-ary function symbols with $n \geq 1$. Theorem 6.8 shows that this is actually an extension (for $\alpha < \omega^\omega$) of Büchi's result.
In fact local α-languages extend far beyond Büchi α-languages: all Büchi ω-languages are, considering topological complexity, boolean combination of G_δ-sets, then Δ^0_3-sets, when there are local ω-languages in each Borel class Σ^0_β, for β an ordinal $< \omega^2$ and there are even some Σ^1_1-complete analytic local ω-languages, [Fin99].
Transfinite length word languages occur in the field of concurrency modelisation with the trace languages [DR95] and also in the work about timed automata where one consider that infinitely many actions may happen during a finite period [BP97] [AD94]. Beyond the decidability of the emptiness problem, we may hope to obtain other decidability results about Büchi α-languages, using results of model theory of local sentences, particularly some stretching theorems based upon the notion of indiscernables in a model [FR96].

In a second paper, we focus on local ω-languages, [Fin99]. We show that:

- Local ω-languages are neither closed under intersection nor under complementation.
- Most undecidability results that hold for locally finite languages may be extended to locally finite ω-languages, in particular the inclusion, the equivalence problems are undecidable, as the problem of the rationality of a local ω-language is.

We then study topological properties of these languages, showing:

- There are local ω-languages in each Borel class of finite rank and even in Borel classes of infinite rank $< \omega^2$.
- For any Borel class G, it is undecidable to determine whether a local ω-language is in the class G.
- There exist local ω-languages which are analytic but not Borel sets.
- One cannot decide whether a local ω-language is a Borel set.

Acknowledgments.

Thanks to the anonymous referee and to M. Nivat for useful comments on the preliminary version of this paper.

References

[ABB96] J-M. Autebert, J. Berstel and L. Boasson, Context Free Languages and Pushdown Automata, in Handbook of Formal Languages, Vol 1, Springer Verlag 1996.

[AD94] R. Alur and D. L. Dill, A Theory of Timed Automata, T.C.S. 126, p. 183-235, 1994.

[Bed96] N. Bedon, Finite Automata and Ordinals, T.C.S. 156, p. 119-144, 1996.

[Bed98] N. Bedon, Langages Reconnaissables de Mots Indexés par des ordinaux, Ph.d Thesis, Université de Marne la Vallée, January 1998.

[Ber79] J. Berstel, Transductions and Context Free Languages, Teubner Studienbücher Informatik, 1979.

[BP97] B. Bérard and C. Picarony, Accepting Zeno Words without Making Time Stand Still, Technical Report, L.S.V., CNRS URA 2236, ENS de Cachan, 1997.

[Büc60] J.R. Büchi, Weak Second Order Arithmetic and Finite Automata, Zeitschrift fur Mathematische Logik und Grundlagen der Mathematik, 6 (1960), pp 66-92.

[Büc62] J.R. Büchi, On a Decision Method in Restricted Second Order Arithmetic, Logic Methodology and Philosophy of Science, (Proc. 1960 Int. Congr.). Stanford University Press, 1962, 1-11.

[BS73] J.R. Büchi, D. Siefkes, The Monadic Second Order Theory of All Countable ordinals, Decidable Theories 2, 1973, S.L.N.M. , number 328.

[CK73] C.C. Chang and H.J. Keisler, Model Theory, American Elsevier Publishing Company, Inc, New York, 1973. (North Holland, Amsterdam, 3rd ed., 1990).

[Cre73] J-P. Crestin, Langages Quasirationnels, In Langages Algébriques, Actes des premières journées d'informatique théorique Bonascre 1973, Ecole Nationale Supérieure de Techniques Avancées, Mars 1978.

[DR95] V. Diekert and G. Rozenberg, Editors. The Book of Traces, World Scientific, Singapore, 1995.

[EF95] H. D. Ebbinghaus and J. Flum, Finite Model Theory, Springer Verlag, New York, 1995.

[Eil74] S. Eilenberg, Automata, Languages and Machines, Vol. A, Academic Press, New York, 1974.

[Fag93] R. Fagin, Finite-Model Theory , A Personal Perspective, T.C.S. 116 (1993) p. 3-31.

[Fin89] O.Finkel, Langages de Büchi et ω-langages Locaux, C.R.A.S. Paris, t.309, Série 1, p. 991-994, 1989.

[Fin93] O.Finkel, Théorie des Modèles des Formules Locales et Etude des Langages Formels qu' elles Definissent, Ph. D Thesis, University Paris 7, May 1993.

[Fin94] O.Finkel, Closure Properties of Locally Finite Omega Languages, preprint, 1994.

[Fin99] O. Finkel, Locally Finite Omega Languages, submitted to T.C.S., september 1999.

[FR96] O. Finkel and J-P. Ressayre, Stretchings, Journal of Symbolic Logic, Volume 61, number 2, Juin 1996, p. 563-585.

[FZV80] P. Franchi-Zanettacci and B. Vauquelin, Automates à Files, T.C.S. 11 (1980), p.221-225.

[Gin66] S. Ginsburg, The Mathematical Theory Of Context Free Languages, Mc Graw-Hill Book Company, New York, 1966.

[Hac98] Y. Hachaichi, Logic For Unambiguous Context Free Languages, submitted to C.S.L. 99.

[Hem92] J. C. Hemmer, Automates sur Mots de Longueur Supérieure à ω, Mémoire de Licence, Université de Liège, 1992.

[HU69] J.E. Hopcroft and J.D. Ullman, Formal Languages and their Relation to Automata, Addison-Wesley Publishing Company, Reading, Massachussetts, 1969.

[Imm87] N. Immerman, Languages That Capture Complexity Classes, SIAM J. of Computing 16:4 (1987), p. 347-354.

[Kas70] T. Kasai, An Hierarchy Between Context Free and Context Sensitive Languages, Journal of Computer and System Sciences, Vol 4, 1970, p.492-508.

[LT94] H. Lescow and W. Thomas, Logical specifications of infinite computations, In:"A Decade of Concurrency" (J. W. de Bakker et al., eds), Springer LNCS 803 (1994), 583-621.

[LST94] C. Lauteman, T. Schwentick and D. Therien, Logics for Context Free Languages, in C.S.L. 94, p. 205-216.

[Ogd69] W.F. Ogden, Intercalation Theorems for Pushdown Store and Stack Languages, Proceeding of the first ACM Symposium on Theory of Computing, 1969, p. 31-42.

[Par83] M. Parigot, Automates, Réseaux, Formules, in : Séminaire Général de Logique, 1983-1984, Publications Mathématiques de l'Université Paris 7.

[PP85] M. Parigot and E. Pelz, A Logical Approach of Petri Net Languages, T.C.S. 39 (1985) p. 155-169.

[PP98] D. Perrin and J.-E. Pin, Mots Infinis, Book in preparation, available from http://www.liafa.jussieu.fr/∼jep/synthese.html

[Pin96] J-E. Pin, Logic, Semigroups and Automata on Words, Annals of Mathematics and Artificial Intelligence 16 (1996), p. 343-384.

[Res88] J-P. Ressayre, Formal Languages defined by the Underlying Structure of their Words, Journal of Symbolic Logic, 53, number 4, December 1988, p. 1009-1026.

[Sch67] J.R. Schoenfield, Mathematical Logic, Addison-Wesley, Reading, MA, 1967.

[Sie65] W. Sierpinski, Cardinal and Ordinal Numbers, Polish Scientific Publisher, Varsovie, 1965.

[Sta97] L. Staiger, ω-languages, Chapter of the Handbbok of Formal languages, Vol 3, edited by G. Rozenberg and A. Salomaa, Springer-Verlag, Berlin.

[Tho90] W. Thomas, Automata on Infinite Objects, in: J. Van Leeuwen, ed., Handbook of Theorical Computer Science, Vol. B (Elsevier, Amsterdam, 1990), p. 133-191.

[Tho96] W. Thomas, Languages, Automata and Logic, in Handbook of Formal Languages Theory, Vol 3, Edited by G. Rozenberg and A. Salomaa, Springer Verlag, 1996.