inria-00564552, version 1 - 9 Feb 2011

# MIX is a 2-MCFL and the word problem in $\mathbb{Z}^2$ is solved by a third-order collapsible pushdown automaton

Sylvain Salvati[a]

[a] *INRIA Bordeaux Sud-Ouest, LaBRI, 351 cours de la Libration F-33405 Talence cedex, France*

**Abstract**

In this work we establish that the language $MIX = \{w \in \{a;b;c\}^* || w|_a = |w|_b = |w|_c\}$ and the language $O_2 = \{w \in \{a;\overline{a};b;\overline{b}\} || w|_a = |w|_{\overline{a}} \wedge |w|_b = |w|_{\overline{b}}\}$ are 2-MCFLs. As 2-MCFLs form a class of languages that is included in both the IO and OI hierarchies, and as $O_2$ is the group language of a simple presentation of $\mathbb{Z}^2$ we exhibit here the first, to our knowledge, non-virtually-free group language (*i.e.* non-context-free group language) that is captured by the IO and OI hierarchies. Moreover, it was a long-standing open problem whether MIX was a mildly context sensitive language or not, and it was conjectured that it was not, so we close this conjecture by giving it a negative answer.

*Keywords:* formal language theory, mildly context sensitive languages, IO and OI hierarchies, higher-order collapsible pushdown automata, group languages, algebraic topology, Jordan curves

## 1. Introduction

The language $MIX = \{w \in \{a;b;c\}^* || w|_a = |w|_b = |w|_c\}$, has been introduced by Emmon Bach in [Bac81] and [Bac88] as a language that proves that the scramble of a context free language may not be context free. This language is also called the Bach language after Geoffrey Pullum [Pul83]. Through the enterprise of Aravind Joshi *et al.* [Jos85] [Wei88] [JSW91] of defining the properties of the class of languages that captures human languages, *i.e.* the class of mildly context sensitive languages, the language $MIX$ has had some importance. Indeed, as it is seen as *an extreme case of the degree of free word order permitted in a language[;] this extreme case is linguistically not relevant* [Jos85]. Following this view, $MIX$ should be excluded from any reasonable class of languages claiming to capture natural languages. It was conjectured in [Jos85] that $MIX$ should not be a Tree Adjoining Language, but there is still no proof of that property which is commented in [JSW91] as being *a very difficult problem* and that it is not even known whether it is an indexed language.

An apparently unrelated problem has attracted the attention of the community of computational group theory. The word problem for a finitely presented (*i.e.* with a finite number of defining relations) group $G$ with finitely many

inria-00564552, version 1 - 9 Feb 2011

generators $\Sigma$ consists in checking whether a word in $\Sigma^*$ is equal to the unit of $G$. This problem is in general undecidable [Nov55] [Boo58] (a quite elegant proof can be found in [Laf09]). So given a finitely presented group $G$ finitely generated by $\Sigma$, the group language defined by this presentation of $G$ is the subset of $\Sigma^*$ which contains all the words that are equal to the unit of $G$. Different presentations of a same group define different group languages but these languages can be translated in each other by rational transductions. A general question (see [Ani71]) consists in relating algebraic properties of finitely presented and finitely generated groups with language-theoretic properties of their group languages. Such a question can be partially answered by characterizing the groups whose group-languages belong to a certain class of languages that is closed under rational transduction (*i.e.* a full trio or rational cone [GGH69]). It is already known that the groups whose group languages are regular languages are precisely *finite groups* and that the ones whose group languages are context-free languages are *virtually free groups* [MS83]. The group language of a simple presentation of $\mathbb{Z}^2$ is the 2-dimensional origin-crossing language (using the terminology and notation of [FR68])

$$O_2 = \{w \in \{a; \bar{a}; b; \bar{b}\}^* || w|_a = |w|_{\bar{a}} \wedge |w|_b = |w|_{\bar{b}}\}$$

which is known not to be context-free. An open problem in computational group theory is whether $O_2$ is an indexed language. We will see in section 3 that actually the problems whether $MIX$ and $O_2$ are indexed languages are in fact equivalent. This shows that this group theoretic question and Joshi's conjecture about $MIX$ are equivalent.

In this paper, we do not solve these open questions, but we prove that $MIX$ and $O_2$ are 2-Multiple Context Free Languages (MCFL) as defined in [SMFK91]. Because MCFLs fall into the class of mildly context sensitive languages that is widely believed to be the class of languages that captures natural languages, it shows that $MIX$ is mildly context sensitive in the sense of [JSW91]. The proof of this result sheds some light on whether $MIX$ and $O_2$ are indexed languages, because the grammar we propose crucially relies on non-well-nested rules which are known to be able to generate languages that are not indexed (see [Mic09] and [KS10]). Finally, the fact that $MIX$ is a mildly context sensitive language is a new result that should be taken into consideration in the debate of defining the class of formal languages in which human languages lie. Furthermore, MCFLs are included in the intersection of IO and OI hierarchies [Dam82] which can be seen as generalizations of macro languages as defined by Fischer [Fis68], and in particular, the OI hierarchy generalizes the notion of indexed languages. As the languages of the OI hierarchy are also captured by higher-order collapsible automata, this shows that this class of automata can solve the word problem for $\mathbb{Z}^2$.

Interestingly, the way we prove that $MIX$ and $O_2$ fall within the class of MCFLs relies on a geometric argument that involves algebraic topology. Thus, most of the proof depends on a Theorem on Jordan curves that may present an interest on its own.

inria-00564552, version 1 - 9 Feb 2011

The paper is structured as follows: section 2 presents some preliminaries, it introduces the notion of Multiple Context Free Grammars and some notions from algebraic topology that play a key role in our proof. In section 3, assuming a Theorem on Jordan curves we prove that $MIX$ and $O_2$ are 2-MCFLs. Section 4 gives the proof of the Theorem on Jordan curves we had admitted so far. Finally section 5 discusses some consequences of the result and some open problems.

## 2. Preliminaries

This section presents the mathematical tools we use in the paper.

### *2.1. Multiple Context Free Grammars*

We write $[n]$ for the set $\{1; \ldots; n\}$ and $\mathbb{N}$ for the set of natural numbers.

For a given finite set $\Sigma$, we write $\Sigma^*$ for the monoid freely generated by $\Sigma$, the elements of $\Sigma$ are called *letters* while the elements of $\Sigma^*$ are called *strings* or *words*. Given a word $w$, we write $|w|$ for its length, and $|w|_x$ for the number of occurrences of the letter $x$ in $w$. A *factor* of a word $w$ is a word $w'$ such that $w = w_1 w' w_2$ for some $w_1$ and $w_2$ in $\Sigma^*$; $w'$ is a *left factor* if $w_1 = \epsilon$ and it is a *right factor* if $w_2 = \epsilon$. The set of factors of $w$ is denoted by $\mathcal{F}(w)$. A factor, a left factor or a right factor is said *trivial* when it is equal to the empty string and *non-trivial* otherwise. The word $w' = v_1 \ldots v_n$ is a *subword* of $w$ if $w = u_0 v_1 u_1 \ldots v_n u_n$. The set of subwords of $w$ is denoted by $\mathcal{S}(w)$. If $w = a_1 \ldots a_n$ where, for each $i$ in $[n]$, $a_i$ is a letter we write $\mathrm{perm}(w)$ for the set of strings defined as:

$$\mathrm{prem}(w) = \{a_{\sigma(1)} \ldots a_{\sigma(n)} | \text{where } \sigma \text{ is a pemutation of } [n]\}.$$

A ranked alphabet $\Omega$ is a pair $(\mathcal{A}, \rho)$ where $\mathcal{A}$ is a finite set and $\rho$ is a function from $\mathcal{A}$ to $\mathbb{N}$. Given $a$ in $\mathcal{A}$, $\rho(a)$ is the *rank* of $a$. We shall write $\Omega^{(n)}$ for the set $\{a \in \mathcal{A} \mid \rho(a) = n\}$.

A Multiple Context Free Grammar (MCFG) $G$ is a tuple $(\Omega, \Sigma, R, S)$ where $\Omega$ is a ranked alphabet, $\Sigma$ is a finite set of letters, $R$ is a set of rules and $S$ is an element of $\Omega^{(1)}$. The rules in $R$ are of the form

$$A(\alpha_1, \ldots, \alpha_n) \text{ :- } B_1(x_1^1, \ldots, x_{l_1}^1), \ldots, B_p(x_1^p, \ldots, x_{l_p}^p)$$

where $A$ is in $\Omega^{(n)}$, $B_j$ is in $\Omega^{(l_j)}$, the $x_j^k$ are pairwise distinct variables and the $\alpha_j$ are elements of $(\Sigma \cup X)^*$ with $X = \{x_j^k \mid k \in [p] \wedge j \in [l_k]\}$ and the restriction that each $x_j^k$ may have at most one occurrence in the string $\alpha_1 \cdots \alpha_n$. Note that $p$ may be equal to 0 in which case the right part of the rule (the one on the right of the :- symbol) is empty, in such a case we will write the rule by omitting the symbol :-.

An MCFG such as $G$ defines *judgments* of the form $\vdash_G A(s_1, \ldots, s_n)$ where $A$ is in $\Omega^{(n)}$ and $s_i$ belongs to $\Sigma^*$. Such a judgment is said to be *derivable* when there is a rule $A(\alpha_1, \ldots, \alpha_n) \text{ :- } B_1(x_1^1, \ldots, x_{l_1}^1), \ldots, B_p(x_1^p, \ldots, x_{l_p}^p)$ and there are

inria-00564552, version 1 - 9 Feb 2011

derivable judgments $\vdash_G B_k(w_1^k, \ldots, w_{l_k}^k)$ for all $k$ in $[p]$ such that $s_j$ is equal to $\alpha_j$ where the possible occurrences of the $x_j^k$ are replaced by $w_j^k$. The language defined by $G$ is the set $\{w \in \Sigma^* \mid S(w) \text{ is derivable}\}$.

An MCFG $G = (\Omega, \Sigma, R, S)$ is a $k$-MCFG when the maximal arity of the elements of $\Omega$ is less than $k$. It is known [SMFK91] that for each $k$, $k$-MCFLs, the languages definable by $k$-MCFG, form substitution-closed full Abstract Family of Languages [GGH69]. In particular, this implies that $k$-MCFLs form a class of languages that is closed under rational transduction for every $k$. Furthermore $k$-MCFLs form a strictly increasing hierarchy of languages.

The notion of MCFG has been introduced in [SMFK91] and MCFLs form a very robust class of languages that is captured by a very wide variety of formalisms. This class of languages is exactly captured by Linear Context Free Rewritting Systems (LCFRSs) [Wei92], Multi-Component Tree Adjoining Grammars (MCTAGs) [JSW91], string languages definable by Hyperedge Replacement Grammars (HR) [Wei92], string languages definable by second order Abstract Categorial Grammars (ACGs) [dGP05] and [Sal07]. Furthermore, since MCTAGs are considered as mildly context sensitive in [JSW91], MCFLs can be thus considered as mildly context sensitive languages as defined by Joshi [Jos85].

The relationship of MCFLs with second order ACGs shows that MCFLs are embedded in both IO and OI hierarchies [Dam82]. As in [dGP05], it is showed that LCFRS can be represented with a second order ACG whose lexicon has complexity 4, *i.e.* a level 3 IO or OI grammar. Even though, it was proved in [DG86] that OI grammars could be recognized by higher-order pushdown automata, it has been done under an implicit restriction, that of *safety*, that has been made explicit later on in [KNU02]. Since then, it is still an open problem whether safe grammars generate the same class of languages as unsafe grammars, and the representation of MCFLs as OI-grammars are unsafe. Notice that level 2 safe grammars and level 2 unsafe grammars have been proved to define the same class of languages [AdMO05], but it is still unclear whether the technique that Aehlig *et al.* have used can be generalized for higher levels. Nevertheless, Hague *et al.* [HMOS08] have proposed an model of automaton that captures the same class of languages as unsafe grammars, *higher-order collapsible automata* such that level $n$ OI languages can be recognized by $n^{th}$-order collapsible automata. So that we have the following fact:

**Fact 1.** *Every MCFL can be recognized with a third-order collapsible pushdown automaton.*

An MCFG is said *well-nested* when all its rules:

$$A(\alpha_1, \ldots, \alpha_n) \coloneq B_1(x_1^1, \ldots, x_{l_1}^1), \ldots, B_p(x_1^p, \ldots, x_{l_p}^p)$$

verify the following properties (where $X = \{x_j^k \mid k \in [p] \wedge j \in [l_k]\}$):

- for $i \in [p]$, if $j < l_i$ then $\alpha_1 \ldots \alpha_n \in (\Sigma \cup X)^* x_j^i (\Sigma \cup X)^* x_{j+1}^i (\Sigma \cup X)^*$,

inria-00564552, version 1 - 9 Feb 2011

- if $i \neq i'$, $j < l_i$ and $j' < l_{i'}$, then $\alpha_1 \dots \alpha_n \notin (\Sigma \cup X)x_{i,j}(\Sigma \cup X)^* x^{i'}_{j'}(\Sigma \cup X)^* x^i_{j+1}(\Sigma \cup X)^* x^{i'}_{j'+1}(\Sigma \cup X)^*$.

This means that the variables of introduced in the right-hand side of the rule appear in the same order in its left hand-side and that furthermore, whenever, for some $i'$ different from $i$, $x^{i'}_{j'}$ occurs in between $x^i_j$ and $x^i_{j+1}$ in $\alpha_1 \dots \alpha_n$, then for all $j''$ in $[l_{i'}]$ the variable $x^{i'}_{j''}$ occurs in between $x^i_j$ and $x^i_{j+1}$. The rules that satisfy these conditions are called *well-nested rules* and the class of languages that can be defined with well-nested MCFG is called *well-nested Multiple Context Free Languages* and written $\text{MCFL}_{wn}$.

Even though this restriction may seem intricate, it decreases the expressive power of MCFGs significantly and $\text{MCFL}_{wn}$ is a very natural class of languages that coincides with many formalisms, like non-duplicating IO and OI grammars (so that $\text{MCFL}_{wn}$ are included in indexed languages), second order ACGs of complexity 3, coupled context-free grammars [Kan09a]. Furthermore, whereas there is still no strong form of pumping lemma for MCFL [KS07], there is one for $\text{MCFL}_{wn}$[Kan09b].

### *2.2. Curves and Homotopy*

We write $\mathbb{R}$ for the set of real numbers equipped with the usual topology and $\mathbb{C}$ for the set of complex numbers also equipped with the usual topology. Given $a$ and $b$ in $\mathbb{R}$, we write $[a;b]$ for the set $\{(1-t)a+tb \mid 0 \leq t \leq 1\}$, $(a;b)$, for the set $\{(1-t)a+tb \mid 0 < t < 1\}$, $[a;b)$ for the set $\{(1-t)a+tb \mid 0 < t \leq 1\}$ and $(a;b]$ for the set $\{(1-t)a+tb \mid 0 \leq t < 1\}$. With this convention $[a;b] = [b;a]$, $(a;b) = (b;a)$ and $[a;b) = (b;a]$. Sets of the form $[a;b]$ and $(a;b)$ are as usual respectively called *closed intervals* and *open intervals*. A closed interval $[a;b]$ is said *trivial* when $a = b$ and *non-trivial* otherwise.

We now introduce some basic notions of algebraic topology, for more details see [Spa81]. Given a topological space $X$, *an arc* in $X$ is a continuous function from $[0,1]$ to $X$. Given an arc $f$ in $X$ such that $f(0) = A$ and $f(1) = B$, we say that $f$ is *an arc from $A$ to $B$*. An arc $f$ in $X$ is said *simple* or is called *a Jordan arc* if $f$ is an injection into $X$. An arc $f$ in $X$ is said to be *a closed curve* if $f(0) = f(1)$ and a closed curve $f$ is said *simple* or is called *a Jordan curve* when the restriction of $f$ to $[0;1)$ is an injection into $X$. Given a closed curve $f$ in $X$ such that $f(0) = A$, we say that $f$ *has base point $A$*. In what follows given an arc $f$ and $R \subseteq [0;1]$, we shall write $f_{|R}$ to denote the restriction of $f$ to $R$. In particular, when $R$ is the interval $[a;b]$, we will make the confusion between $f_{|[a;b]}$ and the arc $f'(t) = f(a+t(b-a))$. When $a \neq b$, we will call $f_{|[a;b]}$ a *subarc* of $f$ and when $f$ is a curve a $f_{|[a;b]}$ is a called a *subcurve* of $f$.

Two arcs $f$ and $g$ from $A$ to $B$ in a space $X$ are said homotop when there is a continuous function $H : [0;1] \times [0;1] \to X$ such that $H(0,t) = f(t)$, $H(1,t) = g(t)$, $H(s,0) = f(0) = g(0)$ and $H(s,1) = f(1) = g(1)$. The relation of *being homotop* is a relation of equivalence and when two arcs $f$ and $g$, in a space $X$, are homotop we shall write $f \approx_X g$.

For the points $A$ and $B$, we write $\mathcal{P}_X(A,B)$ the set of homotopy classes of arcs going from $A$ to $B$ and we write $[f]_X$ for the homotopy class of $f$. If we

inria-00564552, version 1 - 9 Feb 2011

define an operation $_ \cdot _$ of *concatenation of arcs* that, given two arcs $f_1$ and $f_2$, respectively from $A$ to $B$ and $B$ to $C$, produces an arc $f_1 \cdot f_2$ such that

$$f_1 \cdot f_2(t) = \begin{cases} f_1(2t) \text{ when } t \leq \frac{1}{2} \\ f_2(2t-1) \text{ when } t \geq \frac{1}{2} \end{cases}$$

then, when $g_1$ and $g_2$ are respectively homotop to $f_1$ and $f_2$ (*i.e.* $g_1 \in [f_1]_X$ and $g_2 \in [f_2]_X$), we have $[g_1 \cdot g_2]_X = [f_1 \cdot f_2]_X$. Thus we can interpret concatenation at the level of homotopy classes as an operation $_ \bullet _$ of type $\mathcal{P}_X(A,B) \to \mathcal{P}_X(B,C) \to \mathcal{P}_X(A,C)$ such that $[f_1 \cdot f_2]_X = [f_1]_X \bullet [f_2]_X$. This operation can be easily be proved to be associative and to have an inverse for every $[f]_X$, so that the operation $_ \bullet _$ equips the set $\mathcal{P}_X$ of homotopy classes with the algebraic structure of a *groupoid*, *the fundamental groupoid of* $X$. We shall write $\pi(X)$ for the fundamental groupoid of $X$. In particular, when we restrict our attention to homotopy classes to closed curves of base point $A$, then the sub-groupoid we obtain is actually a group, *the fundamental group of* $X$ *at* $A$. We shall write $\pi(X,A)$ for the fundamental group of $X$ at $A$. Given two points $A$ and $B$ of a path connected space $X$ we always have that $\pi(X,A)$ and $\pi(X,B)$ are isomorphic groups. In what follows we will always work in path connected spaces, thus for a given space $X$ we will refer to *its fundamental group*. An arc or a curve $f$ is said *trivial* when its homotopy class is the neutral element of $\pi(X, f(0))$.

The fundamental groupoid of a space $X$ allows expressing particular properties of $X$ that are carried by continuous functions to other spaces. Indeed, given a continuous function $\varphi$ from $X$ to $Y$, $\varphi$ induces a homomorphism $\hat{\varphi}$ between the fundamental groupoids of $X$ and $Y$ because we have:

- whenever $g \in [f]_X$, $\varphi \circ g \in [\varphi \circ f]_Y$,
- $\varphi \circ (f \cdot g) = (\varphi \circ f) \cdot (\varphi \circ g)$,

thus by defining $\hat{\varphi}([f]_X) = [\varphi \circ f]_Y$, we obtain that $\hat{\varphi}([f]_X \bullet [g]_X) = \hat{\varphi}([f \cdot g]_X) = [\varphi \circ (f \cdot g)]_Y = [(\varphi \circ f) \cdot (\varphi \circ g)]_Y = [\varphi \circ f]_Y \bullet [\varphi \circ g]_Y = \hat{\varphi}([f]_X) \bullet \hat{\varphi}([g]_X)$, which shows that $\hat{\varphi}$ is a groupoid homomorphism. Note that if $\varphi$ is a homeomorphism, then $\hat{\varphi}$ is a groupoid isomorphism. Furthermore when we consider closed curves of $X$ at base point $A$, and if $\varphi(A) = B$ then $\hat{\varphi}$ is group homomorphism from $\pi(X,A)$ to $\pi(Y,B)$ (and a group isomorphism when $\varphi$ is a homeomorphism).

Given $X$ and $Y$ two topological spaces, a continuous function $\varphi$ from $X$ to $Y$ is said to *evenly cover* an open subset $U$ of $Y$ when $\varphi^{-1}(U)$ is the disjoint union open subsets of $X$, each of the open sets being homeomorphically mapped on $U$ by $\varphi$. A continuous function from $X$ to $Y$ is a *covering projection* when for every $y$ in $Y$ there is an open neighborhood of $y$ which is evenly covered by $\varphi$. When $\varphi$ is a covering projection from $X$ to $Y$, $X$ is called the *covering space* and $Y$ is called the *base space*, moreover $\varphi$ verifies the following properties:

1. **Unique path-lifting property:** given $f$ an arc of $Y$, given $t$ in $[0;1]$ and $x$ in $\varphi^{-1}(f(t))$, there is a unique arc $g$ of $X$ such that $\varphi \circ g = f$ and $g(t) = x$.

inria-00564552, version 1 - 9 Feb 2011

2. **Homotopy-lifting property:** given $f$ and $g$ two arcs on $X$ such that $f(0) = g(0)$, then whenever $[\varphi \circ f]_Y = [\varphi \circ g]_Y$ we have that $[f]_X = [g]_X$

In section 4 we are going to use these notions so as to prove the following Theorem.

**Theorem 1** *If $f$ is a Jordan curve of $\mathbb{C}$ such that $f(0) = A$ and $f(t) = D$ for some $t$ in $(0;1)$ and there are two points $A'$ and $D'$ in the interior of $f$ such that $D - A = D' - A'$, then there are $t_1$ and $t_2$ such that $t_1 < t_2$ and one of the following properties holds:*

1. $[t_1; t_2] \subseteq (0; t)$ *and* $f(t_2) - f(t_1) = D - A$
2. $[t_1; t_2] \subseteq (t; 1)$ *and* $f(t_2) - f(t_1) = A - D$

## 3. $MIX$ and $O_2$ are a 2-MCFLs

Assuming for the moment Theorem 1, we are going to prove that

$$MIX = \{w \in \{a; b; c\}^* || w|_a = |w|_b = |w|_c\}$$

and

$$O_2 = \{w \in \{a; \bar{a}; b; \bar{b}\}^* || w|_a = |w|_{\bar{a}} \wedge |w|_b = |w|_{\bar{b}}\}$$

are 2-MCFLs. Since the languages $MIX$ and $O_2$ are rationally equivalent (*i.e.* there are rational transductions transforming one language into the other and vice-versa), and that 2-MCFLs form a full AFL, we only need to prove that $O_2$ is a 2-MCFL. The equivalence is established by the following Lemma.

**Lemma 1** *$MIX$ and $O_2$ are rationnally equivalent.*

*Proof.* The Lemma is proved using the following rational transductions[1]:

1. the rational transduction from $O_2$ to $MIX$ can be described by taking the rational set $R_1 = \{a|b|\bar{a}\bar{b}\}^*$ and the string homomorphism $h$ such that $h(a) = a$, $h(b) = b$, $h(\bar{a}) = c$ and $h(\bar{b}) = \epsilon$. We will see below that $MIX = h(O_2 \cap R_1)$.
2. the rational transduction from $MIX$ to $O_2$ is obtained by considering the rational set $R_2 = \{abab|cc|cbcb|aa\}^*$ and the homomorphism $g$ such that $g(a) = abab$, $g(\bar{a}) = cc$, $g(b) = cbcb$ $g(\bar{b}) = aa$. We will show below that $O_2 = g^{-1}(MIX \cap R_2)$.

[1] These transductions are due to Makoto Kanazawa who communicated them to me while exchanging on the problem of MIX. We describe these rational transductions using the bimorphic characterisation due to Nivat [Niv68].

inria-00564552, version 1 - 9 Feb 2011

To prove that the first transduction transforms $O_2$ into $MIX$, it suffices to remark that the language $O_2 \cap R_1$ is the language of strings that contain the same number of occurrences of $a$, $b$ and $\bar{a}\bar{b}$. Thus mapping $a$ to $a$, $b$ to $b$ and $\bar{a}\bar{b}$ to $c$, as the homomorphism $h$ does, gives all the strings contained in $MIX$.

For showing that the second transduction transforms $MIX$ into $O_2$, one needs to remark that a given string $w$ in $MIX \cap R_2$ can be split in a unique way into strings $v_1, \ldots, v_n$ such that $w = v_1 \ldots v_n$ and for all $i$ in $[n]$, $v_i$ is in $\{abab; cc; cbcb; aa\}$ and that if we let $p_1$, $p_2$, $p_3$ and $p_4$ be the number of $v_i$ that are respectively equal to $abab$, $cc$, $cbcb$ and $aa$, then we have that $p_1 + p_4$, $p_1 + p_3$, $p_2 + p_3$ are respectively the number of $a$, $b$ and $c$ that occur in $w$. As $w$ is in $MIX$, we have that the following identity holds:

$$p_1 + p_4 = p_2 + p_3 = p_1 + p_3$$

but this identity is equivalent to have $p_1 = p_2$ and $p_3 = p_4$. Thus as $w$ is in $MIX$, we have obviously $g^{-1}(w)$ in $O_2$. Conversely it is obvious that whenever $w$ is in $O_2$, then $g(w)$ is in $MIX \cap R_2$ so that we indeed have $O_2 = g^{-1}(MIX \cap R_2)$. □

The main argument of the proof that $O_2$ is a 2-MCFL is of geometric nature. We define an interpretation $\varphi(w)$ of the words of $\{a; \bar{a}; b; \bar{b}\}^*$ in $\mathbb{Z} + i\mathbb{Z}$ a subset of the complex numbers $\mathbb{C}$:

- $\varphi(\epsilon) = 0$
- $\varphi(aw) = \varphi(w) + i$ and $\varphi(\bar{a}w) = \varphi(w) - i$
- $\varphi(bw) = \varphi(w) + 1$ and $\varphi(\bar{b}w) = \varphi(w) - 1$

Obviously $\varphi$ is a monoid homomorphism and $O_2 = \varphi^{-1}(0)$ (this is just another way of saying that $O_2$ is the group language of $\mathbb{Z} + i\mathbb{Z}$ which is isomorphic to $\mathbb{Z}^2$).

Given $w$ in $\{a; \bar{a}; b; \bar{b}\}^*$ such $w = c_1 \ldots c_n$ with $c_i$ in $\{a; \bar{a}; b; \bar{b}\}$, we interpret $w$ as the arc $\mathcal{C}(w)$ defined by:

$$\mathcal{C}(w)(t) = \begin{cases} nt\varphi(c_1) \text{ when } t \leq \frac{1}{n} \\ \varphi(c_1 \ldots c_{k-1}) + n(t - \frac{k-1}{n})\varphi(c_k) \text{ when } \frac{k-1}{n} < t \leq \frac{k}{n} \end{cases}$$

Notice that the construction of $\mathcal{C}(w)$ implies that every $t$ in $[0; 1]$, either the imaginary or the real part of $\mathcal{C}(w)(t)$ is in $\mathbb{Z}$. Furthermore, we have the following fact:

**Fact 2.** *Given $t$ in $[0; 1]$, both the imaginary and parts of $\mathcal{C}(w)(t)$ are in $\mathbb{Z}$ iff $t = \frac{k}{|w|}$ with $0 \leq k \leq |w|$, in such a case we obviously have that $\mathcal{C}(w)(t) = \varphi(v_k)$ if $v_k$ is the left factor of length $k$ of $w$.*

**Lemma 2** *Given $z$ in $\mathbb{Z} + i\mathbb{Z}$, a word $w$ and $t_1$, $t_2$ in $[0; 1]$, if $\mathcal{C}(w)(t_1)$ is not in $\mathbb{Z} + i\mathbb{Z}$ and $\mathcal{C}(w)(t_1) - \mathcal{C}(w)(t_2) = z$ then if $k_1$ and $k_2$ are the integers such that $\frac{k_1}{|w|} \leq t_1 \leq \frac{k_1+1}{|w|}$, and $\frac{k_2}{|w|} \leq t_2 \leq \frac{k_2+1}{|w|}$, then one of the following properties holds:*

inria-00564552, version 1 - 9 Feb 2011

1. *either* $\mathcal{C}(w)(\frac{k_1}{|w|}) - \mathcal{C}(w)(\frac{k_2}{|w|}) = \mathcal{C}(w)(\frac{k_1+1}{|w|}) - \mathcal{C}(w)(\frac{k_2+1}{|w|}) = z$,
2. *or* $\mathcal{C}(w)(\frac{k_1+1}{|w|}) - \mathcal{C}(w)(\frac{k_2}{|w|}) = \mathcal{C}(w)(\frac{k_1}{|w|}) - \mathcal{C}(w)(\frac{k_2+1}{|w|}) = z$.

*Proof.* As $\mathcal{C}(w)(t_1)$ is not in $\mathbb{Z} + i\mathbb{Z}$, only the imaginary part or the real part of $\mathcal{C}(w)(t_1)$ is in $\mathbb{Z}$. Without loss of generality we assume that the real part of $\mathbb{C}(w)(t_1)$ is in $\mathbb{Z}$ (the other case can be treated in a similar way) and that $\mathcal{C}(w)(t_1) = h + i(l + \delta)$ with, $h$ and $l$ in $\mathbb{Z}$ and $\delta$ in $(0; 1)$. By definition of $\mathcal{C}$ we have that one of the following holds:

1. either $\mathcal{C}(w)(\frac{k_1}{|w|}) = h + il$ and $\mathcal{C}(w)(\frac{k_1+1}{|w|}) = h + i(l + 1)$,
2. or $\mathcal{C}(w)(\frac{k_1}{|w|}) = h + i(l + 1)$ and $\mathcal{C}(w)(\frac{k_1+1}{|w|}) = h + il$.

Since $\mathcal{C}(w)(t_1) - \mathcal{C}(w)(t_2) = z$ we have $\mathcal{C}(w)(t_2) = h + i(l + \delta) + z$. And as $z$ is in $\mathbb{Z} + i\mathbb{Z}$, as previously we must have

1. either $\mathcal{C}(w)(\frac{k_2}{|w|}) = h + il + z$ and $\mathcal{C}(w)(\frac{k_2+1}{|w|}) = h + i(l + 1) + z$,
2. or $\mathcal{C}(w)(\frac{k_2}{|w|}) = h + i(l + 1) + z$ and $\mathcal{C}(w)(\frac{k_2+1}{|w|}) = h + il + z$,

from which the conclusion follows. □

Notice also that $w$ is in $O_2$ if and only if $\mathcal{C}(w)$ is a closed curve. It is easy to establish that $\mathcal{C}(w)$ satisfies the following properties:

- if $w = \epsilon$ then $\mathcal{C}(w)(t) = 0$
- if $w \in \{a; \overline{a}; b; \overline{b}\}$ then $\mathcal{C}(w)(1) = \varphi(w)$
- if $w = w_1 w_2$ then $\mathcal{C}(w)(t) = \begin{cases} \mathcal{C}(w_1)(t\frac{|w_1 w_2|}{|w_1|}) \text{ when } t < \frac{|w_1|}{|w_1 w_2|} \\ \mathcal{C}(w_2)(t\frac{|w_1 w_2|}{|w_2|} - \frac{|w_1|}{|w_2|}) + \varphi(w_1) \text{ otherwise} \end{cases}$

An example of an arc representation of a word is given by figure 1, where a black dot is used to mark the starting point and an arrow is giving the starting direction.

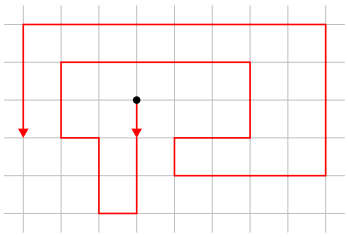


Figure 1: $\mathcal{C}(w)$ when $w = \overline{aa}\overline{a}\overline{b}aa\overline{b}aabbbbb\overline{aa}\overline{bb}\overline{a}bbbbaaaa\overline{bbbbbbbb}\overline{aa}\overline{a}$

A word $w$ is said *simple* if for every element $w'$ of $\mathcal{F}(w)/\{\epsilon\}$, $\varphi(w') \neq 0$. Fact 2 implies:

**Fact 3.** *$w$ is a simple word iff $\mathcal{C}(w)$ is a simple arc.*

inria-00564552, version 1 - 9 Feb 2011

A subword $w'$ of $w$ is a *simplification* of $w$ when $w'$ is simple, $w' = v'_1 \dots v'_n$ and $w = u_1 v'_1 \dots u_n v'_n u_{n+1}$ with $\varphi(u_i) = 0$ for $i$ in $[n+1]$. The set of simplifications of $w$ will be written $\mathrm{sp}(w)$.

**Lemma 3** *If $w$ is not in $O_2$ then $\mathrm{sp}(w) \neq \emptyset$.*

*Proof.* We proceed by induction on the size of $w$.

In case $w$ is simple then $w$ is in $\mathrm{sp}(w)$.

In case $w$ is not simple, let $u$ be non-trivial factor of $w$ such that $\varphi(u) = 0$ and $w = w_1 u w_2$. Since, $w$ is not in $O_2$, $w_1 w_2$ is also not in $O_2$ and thus, by induction hypothesis, $\mathrm{sp}(w_1 w_2)$ is not empty. Let $w'$ be in $\mathrm{sp}(w_1 w_2)$, we are going to see that $w'$ is in $\mathrm{sp}(w)$. Indeed, since $w'$ is in $\mathrm{sp}(w_1 w_2)$, $w' = v'_1 \dots v'_n$ and $w_1 w_2 = u_1 v'_1 \dots u_n v'_n u_{n+1}$ with $\varphi(u_i) = 0$ for all $i$ in $[n+1]$. This implies that:

1. either $w_1 = u_1 v'_1 \dots u_k v'_{l,k}$ and $w_2 = v'_{r,k} u_{k+1} \dots v'_n u_{n+1}$ with $v_k = v'_{l,k} v'_{r,k}$
2. or $w_1 = u_1 v'_1 \dots v'_{k-1} u_{l,k}$ and $w_2 = u_{r,k} v'_k \dots u_n v'_n u_{n+1}$ with $u_k = u_{l,k} u_{r,k}$

In the first case, we have that $w = u_1 v'_1 \dots u_k v'_{l,k} u v'_{r,k} \dots u_n v'_n u_{n+1}$ and since $\varphi(u) = 0$ and for all $i$ in $[n+1]$, $\varphi(u_i) = 0$, we have that $w'$ in $\mathrm{sp}(w)$.

In the second case, we have that $w = u_1 v'_1 \dots u_k u_{l,k} u u_{r,k} v'_k \dots u_n v'_n u_{n+1}$ and since $\varphi(u_{l,k} u u_{r,k}) = \varphi(u) + \varphi(u_{l,k} u_{r,k}) = 0$ and for all $i$ in $[n+1]$ with $i \neq k$ we have $\varphi(u_i) = 0$ we finally obtain that $w'$ is in $\mathrm{sp}(w)$. □

**Lemma 4** *If $w' \in sp(w)$ and $w' = w'_1 w'_2$ with $w'_1 \neq \epsilon$ and $w'_2 \neq \epsilon$ then $w = w_1 w_2$ such that $w'_1$ and $w'_2$ are respectively in $\mathrm{sp}(w_1)$ and $\mathrm{sp}(w_2)$.*

*Proof.* If $w'$ is in $\mathrm{sp}(w)$, then $w = u_1 v'_1 \dots u_n v'_n u_{n+1}$, $w' = v'_1 \dots v'_n$, and $\varphi(u_i) = 0$ for all $i$ in $[n+1]$. This implies that $w'_1 = v'_1 \dots v'_{k-1} v''_{1,k}$ and $w'_2 = v''_{2,k} v'_{k+1} \dots v'_n$ for some $k$ in $[n]$ and $v'_k = v''_{1,k} v''_{2,k}$. It suffices to choose $w_1 = u_1 v'_1 \dots v_{k-1} u_k v''_{1,k}$ and $w_2 = v''_{2,k} u_{k+1} v'_{k+1} \dots u_n v'_n u_{n+1}$. □

**Corollary 1.** *If $w' \in \mathrm{sp}(w)$ and $w' = w'_1 \dots w'_n$ with $w'_i \neq \epsilon$ for all $i$ in $[n]$, then $w = w_1 \dots w_n$ such that for all $i$ in $[n]$ $w'_i$ is in $\mathrm{sp}(w_i)$.*

*Proof.* A simple iteration of the preceding Lemma. □

We define $G = (\Omega, \{a; \overline{a}; b; \overline{b}\}, R, S)$ where $\Omega = (\{S; \mathrm{Inv}\}, \rho)$ with $\rho(S) = 1$ and $\rho(\mathrm{Inv}) = 2$ and where $R$ is made of the rules that have one of the following forms:

1. $S(x_1 x_2) \text{ :- } \mathrm{Inv}(x_1, x_2)$,
2. $\mathrm{Inv}(t_1, t_2) \text{ :- } \mathrm{Inv}(x_1, x_2)$ where $t_1 t_2 \in \mathrm{perm}(x_1 x_2 a \overline{a}) \cup \mathrm{prem}(x_1 x_2 b \overline{b})$,
3. $\mathrm{Inv}(t_1, t_2) \text{ :- } \mathrm{Inv}(x_1, x_2), \mathrm{Inv}(y_1, y_2)$ where $t_1 t_2 \in \mathrm{prem}(x_1 x_2 y_1 y_2)$
4. $\mathrm{Inv}(\epsilon, \epsilon)$

inria-00564552, version 1 - 9 Feb 2011

*N.B.:* the items 2 and 3 represent finite sets of rules. It is easy to see that the language generated by $G$ is included in $O_2$. We will see that the converse also holds.

We are going to show the following Lemma about this grammar:

**Lemma 5** *Whenever $w_1$ and $w_2$ are elements of $\{a; \overline{a}; b; \overline{b}\}^*$ such that $w_1 w_2$ is in $O_2$, $\mathrm{Inv}(w_1, w_2)$ is derivable in $G$.*

As a consequence we obtain that the language of $G$ is precisely $O_2$ and thus the expected Theorem.

**Theorem 2** *$MIX$ and $O_2$ are 2-MCFLs.*

Fact 1 implies the following corollary.

**Corollary 2.** *$MIX$ and $O_2$ are recognized by a third-order collapsible pushdown automaton.*

*Proof of Lemma 5.* The proof is done by induction on $|w_1 w_2| + \max(|w_1|, |w_2|)$. There are five cases:

**Case 1** $w_1$ and $w_2$ are in $O_2$. In case neither of them is the empty string, we have, by induction hypothesis, that $\mathrm{Inv}(w_1, \epsilon)$ and $\mathrm{Inv}(w_2, \epsilon)$ are derivable so that $\mathrm{Inv}(w_1, w_2)$ is derivable using the rule $\mathrm{Inv}(x_1 x_2, y_1 y_2) \text{ :- } \mathrm{Inv}(x_1, x_2), \mathrm{Inv}(y_1, y_2)$.

In case either $w_1$ or $w_2$ is the empty string then when both of them are equal to the empty string, $\mathrm{Inv}(w_1, w_2)$ is obviously derivable using rule 4. In case $w_1$ or $w_2$ is not the empty string, without loss of generality, we may assume that $w_2$ is equal to the empty string and $w_1$ is not. Then, we have $w_1$ in $O_2$ and $w_1 \neq \epsilon$, so that, $w_1$ has length at least 2 and there is $v_1$ and $v_2$ both different from the empty string such that $w_1 = v_1 v_2$. So by induction hypothesis $\mathrm{Inv}(v_1, v_2)$ is derivable and using the rule $\mathrm{Inv}(x_1 x_2, y_1 y_2) \text{ :- } \mathrm{Inv}(x_1, x_2)$, $\mathrm{Inv}(y_1, y_2)$ together with the premises $\mathrm{Inv}(v_1, v_2)$ and $\mathrm{Inv}(\epsilon, \epsilon)$ we obtain a derivation of $\mathrm{Inv}(v_1 v_2, \epsilon)$, *i.e.* a derivation of $\mathrm{Inv}(w_1, w_2)$.

**Case 2** In case $w_1 = s_1 w_1' s_2$, $w_2 = s_3 w_2' s_4$ so that $s_1 s_2 s_3 s_4 \in \{a\overline{a}; \overline{a}a; b\overline{b}; \overline{b}b\}$, then we can conclude by applying one of the rules defined by the item 2 in the definition of $G$. For example if $w_1 = a w_1' \overline{a}$ then we let $s_1 = a$, $s_2 = \overline{a}$, $s_3 = s_4 = \epsilon$, $w_2' = w_2$, and we have $s_1 s_2 s_3 s_4 = a\overline{a}$. So by induction hypothesis, $\mathrm{Inv}(w_1', w_2)$ is derivable, and we may derive $\mathrm{Inv}(w_1, w_2)$ simply by using the rule $\mathrm{Inv}(a x_1 \overline{a}, x_2) \text{ :- } \mathrm{Inv}(x_1, x_2)$.

**Case 3** Neither $w_1$ nor $w_2$ is the empty string and either $w_1$ or $w_2$ has a non-trivial left or right factor that is in $O_2$. Without loss of generality, we may suppose that $w_1 = v w_1'$ where $v$ is not the empty string and $v$ in $O_2$. By induction hypothesis (that can be used since neither $v$ nor $w_2$ are the empty string) $\mathrm{Inv}(v, \epsilon)$ and $\mathrm{Inv}(w_1', w_2)$ are derivable, and then $\mathrm{Inv}(w_1, w_2)$ is derivable using the rule $\mathrm{Inv}(x_1 y_1, y_2 x_2) \text{ :- } \mathrm{Inv}(x_1, x_2)$, $\mathrm{Inv}(y_1, y_2)$ of the family of rules 3.

inria-00564552, version 1 - 9 Feb 2011

**Case 4** Neither $w_1$ nor $w_2$ is in $O_2$ and there is $w'_1$ and $w'_2$ respectively in $\mathrm{sp}(w_1)$ and $\mathrm{sp}(w_2)$ such that the curves $\mathcal{C}(w'_1w'_2)$ is not a simple curve. From fact 3, we know that $\mathcal{C}(w'_1)$ and $\mathcal{C}(w'_2)$ are simple arcs. Since

$$\mathcal{C}(w'_1w'_2)(t) = \begin{cases} \mathcal{C}(w'_1)(t\frac{|w'_1w'_2|}{|w'_1|}) \text{ when } t < \frac{|w'_1|}{|w'_1w'_2|} \\ \mathcal{C}(w'_2)(t\frac{|w'_1w'_2|}{|w'_2|} - \frac{|w'_1|}{|w'_2|}) + \varphi(w'_1) \text{ otherwise} \end{cases}$$

and since $\mathcal{C}(w'_1)$ and $\mathcal{C}(w'_2)$ are simple while $\mathcal{C}(w'_1w'_2)$ is not simple, there must be $t_1$ and $t_2$ respectively in $(0, \frac{|w'_1|}{|w'_1w'_2|})$ and $(\frac{|w'_1|}{|w'_1w'_2|}, 1)$ such that:

$$\mathcal{C}(w'_1w'_2)(t_1) \quad = \quad \mathcal{C}(w'_1w'_2)(t_2) \tag{1}$$

which can be rephrased as:

$$\mathcal{C}(w'_1)(t_1\frac{|w'_1w'_2|}{|w'_1|}) \quad = \quad \mathcal{C}(w'_2)(t_2\frac{|w'_1w'_2|}{|w'_2|} - \frac{|w'_1|}{|w'_2|}) + \varphi(w'_1) \tag{2}$$

Moreover, either the imaginary part or the real part of $\mathcal{C}(w'_1w'_2)(t_1)$ is in $\mathbb{Z}$.

In case both the imaginary and the real part of $\mathcal{C}(w'_1w'_2)(t_1) = \mathcal{C}(w'_1)(t_1\frac{|w'_1w'_2|}{|w'_1|})$ are in $\mathbb{Z}$ then, fact 2 implies that $t_1\frac{|w'_1w'_2|}{|w'_1|} = \frac{k}{|w'_1|}$ with $0 \leq k \leq |w'_1|$. But since $t_1$ is in $(0, \frac{|w'_1|}{|w'_1w'_2|})$ we must have that $0 < k < |w'_1|$. Fact 2 also implies that $w'_1 = v'_1v'_2$ with $v'_1$ being the left factor of length $k$ of $w'_1$ and $\varphi(v'_1) = \mathcal{C}(w'_1)(t_1\frac{|w'_1w'_2|}{|w'_1|})$. From the fact that $0 < k < |w'_1|$, we have that neither $v'_1$ nor $v'_2$ is the empty string. Similarly $w'_2 = v'_3v'_4$ with $\mathcal{C}(w'_2)(t_2\frac{|w'_1w'_2|}{|w'_2|} - \frac{|w'_1|}{|w'_2|}) = \varphi(v'_3)$. But (2) implies that $\varphi(v'_1) = \varphi(v'_3) + \varphi(w'_1)$ and, since $\varphi(w'_1) = \varphi(v'_1) + \varphi(v'_2)$, we obtain $\varphi(v'_2) + \varphi(v'_3) = 0$ so that $v'_2v'_3$ is in $O_2$. Because $w_1w_2$ is in $O_2$, $w'_1w'_2$ is also in $O_2$ and therefore, $\varphi(w'_2) = -\varphi(w'_1)$ so that with $\varphi(v'_1) = \varphi(v'_3) + \varphi(w'_1)$ and $\varphi(w'_2) = \varphi(v'_3)+\varphi(v'_4)$ we have $\varphi(v'_1)+\varphi(v'_4) = 0$ and $v'_1v'_4$ is in $O_2$ (see figure 2). Now, Lemma 4, implies that $w_1 = v_1v_2$ and $w_2 = v_3v_4$ so that $v'_i$ is in $\mathrm{sp}(v_i)$ for all $i$ in [4] and $v_1v_4$ and $v_2v_3$ are in $O_2$ and since neither $v'_1$ nor $v'_2$ is the empty string, it is also the case that neither $v_1$ nor $v_2$ is the empty string which has the consequence that $|v_1v_4| < |w_1w_2|$ and $|v_2v_3| < |w_1w_2|$. Then, by induction hypothesis, $\mathrm{Inv}(v_1, v_4)$ and $\mathrm{Inv}(v_2, v_3)$ are derivable, which, with the rule $\mathrm{Inv}(x_1y_1, y_2x_2)$ :- $\mathrm{Inv}(x_1, x_2), \mathrm{Inv}(y_1, y_2)$, allows us to finally derive $\mathrm{Inv}(w_1, w_2)$.

In case only the imaginary or real part of $\mathcal{C}(w'_1w'_2)(t_1)$ is an integer let $k_1$ and $k_2$ be the integer such that $\frac{k_1}{|w'_1w'_2|} < t_1 < \frac{k_1+1}{|w'_1w'_2|}$ and $\frac{k_2}{|w'_1w'_2|} < t_2 < \frac{k_2+1}{|w'_1w'_2|}$. Notice that since $t_1 \in (0; \frac{|w'_1|}{|w'_1w'_2|})$ and $t_2 \in (\frac{|w'_1|}{|w'_1w'_2|}; 1)$, we must have $0 \leq k_1 < k_1 + 1 \leq |w'_1|$ and $|w'_1| \leq k_2 < k_2 + 1 \leq |w'_1w'_2|$. The identity (1), implies that $\mathcal{C}(w'_1w'_2)(t_1) - \mathcal{C}(w'_1w'_2)(t_2)$ is in $\mathbb{Z} + i\mathbb{Z}$ so that Lemma 2 implies that

1. either $\mathcal{C}(w'_1w'_2)(\frac{k_1}{|w'_1w'_2|}) = \mathcal{C}(w'_1w'_2)(\frac{k_2}{|w'_1w'_2|})$
   and
   $\mathcal{C}(w'_1w'_2)(\frac{k_1+1}{|w'_1w'_2|}) = \mathcal{C}(w'_1w'_2)(\frac{k_2+1}{|w'_1w'_2|})$,

inria-00564552, version 1 - 9 Feb 2011

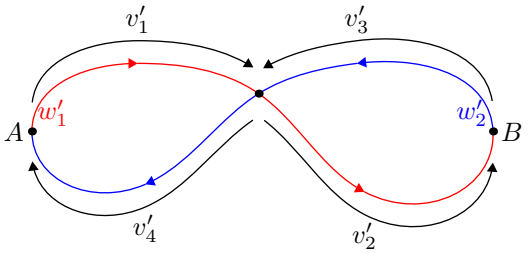



Figure 2: $\mathcal{C}(w'_1 w'_2)$ is not a simple curve (Case 4)

2. or $\mathcal{C}(w'_1 w'_2)(\frac{k_1}{|w'_1 w'_2|}) = \mathcal{C}(w'_1 w'_2)(\frac{k_2+1}{|w'_1 w'_2|})$
   and
   $\mathcal{C}(w'_1 w'_2)(\frac{k_1+1}{|w'_1 w'_2|}) = \mathcal{C}(w'_1 w'_2)(\frac{k_2}{|w'_1 w'_2|})$.

Which is equivalent to (letting $k'_2 = k_2 - |w'_1|$):

1. either $\mathcal{C}(w'_1)(\frac{k_1}{|w'_1|}) = \mathcal{C}(w'_2)(\frac{k'_2}{|w'_2|}) + \varphi(w'_1)$
   and
   $\mathcal{C}(w'_1)(\frac{k_1+1}{|w'_1|}) = \mathcal{C}(w'_2)(\frac{k'_2+1}{|w'_2|}) + \varphi(w'_1)$
2. or $\mathcal{C}(w'_1)(\frac{k_1}{|w'_1|}) = \mathcal{C}(w'_2)(\frac{k'_2+1}{|w'_2|}) + \varphi(w'_1)$
   and
   $\mathcal{C}(w'_1)(\frac{k_1+1}{|w'_1|}) = \mathcal{C}(w'_2)(\frac{k'_2}{|w'_2|}) + \varphi(w'_1)$.

Thus, in case either $0 < k_1 < |w'_1|$, or $0 < k_1 + 1 < |w'_1|$, or $0 < k'_2 < |w'_2|$, or $0 < k'_2 + 1 < |w'_2|$, we may conclude as above. Otherwise, $k_1 = 0$, $k_1 + 1 = |w'_1|$, $k'_2 = 0$ and $k'_2 + 1 = |w'_2|$ which implies that $|w'_1| = 1$ and $|w'_2| = 1$. Thus there are four possibilities: either $w'_1 = a$ and $w'_2 = \overline{a}$, or $w'_1 = \overline{a}$ and $w'_2 = a$, or $w'_1 = b$ and $w'_2 = \overline{b}$, or $w'_1 = \overline{b}$ and $w'_2 = b$. Without loss of generality, we assume that $w'_1 = a$ and $w'_2 = \overline{a}$. As a consequence, $w_1 = u_1 a u_2$ and $w_2 = u_3 \overline{a} u_4$ with $\varphi(u_1) = \varphi(u_2) = \varphi(u_3) = \varphi(u_4) = 0$, if one of the $u_i$ is not the empty string, then the induction works as in case 3, otherwise, if $w_1 = w'_1 = a$ and $w_2 = w'_2 = \overline{a}$, and, similarly to case 2, it suffices to derive $\mathrm{Inv}(w_1, w_2)$ (*i.e.* $\mathrm{Inv}(a, \overline{a})$) from $\mathrm{Inv}(\epsilon, \epsilon)$ and the rule $\mathrm{Inv}(a x_1, \overline{a} x_2) \text{ :- } \mathrm{Inv}(x_1, x_2)$.

**Case 5** The last case is the complement of all the previous cases, $w_1$ and $w_2$ verify the following properties:

1. neither $w_1$ nor $w_2$ is in $O_2$,
2. $w_1$ and $w_2$ do not start of end with compatible letters,
3. neither $w_1$ nor $w_2$ has a non-trivial left or right factor that belongs to $O_2$,
4. if $w'_1$ and $w'_2$ are respectively in $\mathrm{sp}(w_1)$ and $\mathrm{sp}(w_2)$, then $\mathcal{C}(w'_1 w'_2)$ is a simple curve (*i.e.* a Jordan curve).

From Lemma 3, because $w_1 \notin O_2$ and $w_2 \notin O_2$, there are $w'_1$ and $w'_2$ respectively in $\mathrm{sp}(w_1)$ and $\mathrm{sp}(w_2)$. By definition it must be that case that $w_1 =$

inria-00564552, version 1 - 9 Feb 2011

$u_1v'_1 \dots u_nv'_nu_{n+1}$ with $\varphi(u_i) = 0$ for all $i$ in $[n+1]$ and $w'_1 = v'_1 \dots v'_n$. But as, by hypothesis, $w_1$ may not have a non-trivial left or right factor in $O_2$, we must have $u_1 = u_{n+1} = \epsilon$ so that first and last letters of $w'_1$ are the same as the first and last letters of $w_1$. Similarly, the same holds for $w_2$ and $w'_2$. Because $\mathcal{C}(w'_1w'_2)$ is a Jordan curve, it splits the plane into two components, the interior of the curve and the exterior of the curve, furthermore, while traversing the curve the interior and the exterior are always on the same side of the curve. We call *unit square* in $\mathbb{Z} + i\mathbb{Z}$ a set of points of $\mathbb{C}$ that lie inside a square whose diagonal is given by $z$ and $z+1+i$ with $z$ in $\mathbb{Z}+i\mathbb{Z}$. A unit square is *adjacent* to $\mathcal{C}(w'_1w'_2)$ if one of its corners is on $\mathcal{C}(w'_1w'_2)$. The fact that $\mathcal{C}(w'_1w'_2)$ is a Jordan curve implies that while traversing $\mathcal{C}(w'_1w'_2)$:

1. either the interior every unit square adjacent to $\mathcal{C}(w'_1w'_2)$ appearing on the right of are in the interior of $\mathcal{C}(w'_1w'_2)$ while the interior of the ones appearing on the left are in its exterior,
2. or the converse.

Exploiting this fact, figure 3 shows all the possible cases (omitting the cases that can be obtained by symmetry; in particular it focuses on the cases where the first and last letters of $w'_1$ and $w'_2$ are either $a$ or $b$), the red arrows represent the starting and ending letters of $w'_1$ while the blue ones represent the starting and ending letters of $w'_2$ and we have materialized with green vectors (*resp.* yellow vectors) two points that must be either both outside $\mathcal{C}(w'_1w'_2)$ or both inside $\mathcal{C}(w'_1w'_2)$. Moreover when the pair of points defined with the green vectors, are outside the $\mathcal{C}(w'_1w'_2)$, then the points defined with the yellow vectors are inside $\mathcal{C}(w'_1w'_2)$ and conversely (figure 4 exemplifies the two different cases for a particular configuration of the starting and ending letters of $w_1$ and $w_2$). In that context we can apply Theorem 1 considering that $f = \mathcal{C}(w'_1w'_2)$ is a Jordan curve, that the points $A$ and $D$ of $\mathcal{C}(w'_1w'_2)$ are respectively given by $\mathcal{C}(w'_1w'_2)(0)$ and $\mathcal{C}(w'_1w'_2)(\frac{|w'_1|}{|w'_2|})$ (which are the starting and ending points of the part of $\mathcal{C}(w'_1w'_2)$ that represents $w'_1$), the points $A'$ and $D'$ are respectively either $A+z$ and $D+z$, or $A-z$ and $D-z$ with $z = -\frac{1}{2} + i\frac{1}{2}$. Then Theorem 1 gives us $t_1$ and $t_2$ such that $t_1 < t_2$ and such that one of the following statements holds:

1. $[t_1; t_2] \subseteq (0; \frac{|w'_1|}{|w'_2|})$ and $f(t_2) - f(t_1) = D - A$
2. $[t_1; t_2] \subseteq (\frac{|w'_1|}{|w'_2|}; 1)$ and $f(t_2) - f(t_1) = A - D$

We may suppose that the first statement holds (the other case being symmetric) and in order to conclude we need to prove that this gives the existence of some non-empty strings $v'_1$, $v'_2$ and $v'_3$ such that $w'_1 = v'_1v'_2v'_3$ and $\varphi(v'_2) = \varphi(w'_1)$. Once this is proved, we have, using Lemma 1, that there are non-empty strings $v_1$, $v_2$, and $v_3$ such that $w_1 = v_1v_2v_3$ and $\varphi(v_2) = \varphi(w_1)$ so that $v_1v_3$ and $v_2w_2$ are in $O_2$ and that, by induction hypothesis $\mathrm{Inv}(v_1, v_3)$ and $\mathrm{Inv}(v_2, w_2)$ are derivable from which, using the rule $\mathrm{Inv}(x_1y_1x_2, y_2) \coloneq \mathrm{Inv}(x_1, x_2)\,\mathrm{Inv}(y_1, y_2)$, we can conclude that $\mathrm{Inv}(w_1, w_2)$ is derivable.

inria-00564552, version 1 - 9 Feb 2011

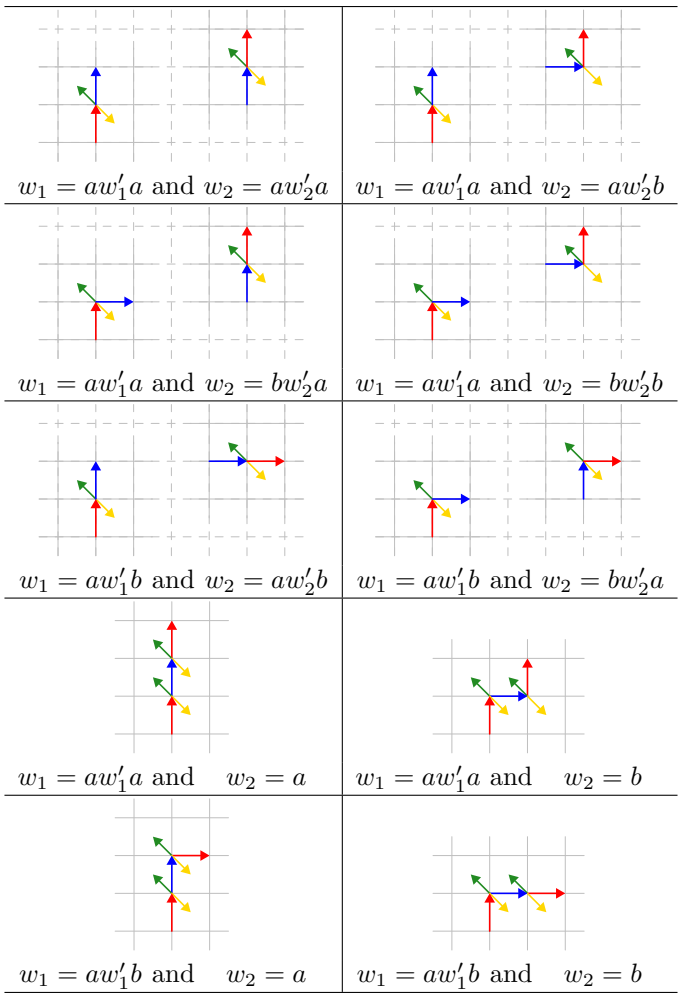



Figure 3: Possible independant configurations for the last case of the proof

Proving the existence of $v'_1$, $v'_2$ and $v'_3$ relies on fact 2. Indeed, once it is proved that there are $t'_1$ and $t'_2$ such that $t'_1 < t'_2$, $[t'_1; t'_2] \subseteq (0; \frac{|w'_1|}{|w'_2|})$, $f(t'_2) - f(t'_1) = D - A$ and such that $f(t'_1)$ and $f(t'_2)$ have both their imaginary and real parts in $\mathbb{Z}$, fact 2 gives us $v'_1$, $v'_2$ and $v'_3$. In case both $f(t_1)$ and $f(t_2)$ are in $\mathbb{Z} + i\mathbb{Z}$ then we are done simply by choosing $t'_1 = t_1$ and $t'_2 = t_2$. Otherwise, by construction of $D$ and $A$, we have that $D - A$ is in $\mathbb{Z} + i\mathbb{Z}$. We let $k_1$ and $k_2$ be such that:

1. $\frac{k_1}{|w'_1|} < t_1 \frac{|w'_1 w'_2|}{|w'_1|} < \frac{k_1+1}{|w'_1|}$,
2. $\frac{k_2}{|w'_1|} < t_2 \frac{|w'_1 w'_2|}{|w'_1|} < \frac{k_2+1}{|w'_1|}$.

inria-00564552, version 1 - 9 Feb 2011

and Lemma 2 implies that:

1. either $\mathcal{C}(w'_1)(\frac{k_2}{|w'_1|}) - \mathcal{C}(w'_1)(\frac{k_1}{|w'_1|}) = D - A$ and $\mathcal{C}(w'_1)(\frac{k_2+1}{|w'_1|}) - \mathcal{C}(w'_1)(\frac{k_1+1}{|w'_1|}) = D - A$,
2. or $\mathcal{C}(w'_1)(\frac{k_2+1}{|w'_1|}) - \mathcal{C}(w'_1)(\frac{k_1}{|w'_1|}) = D - A$ and $\mathcal{C}(w'_1)(\frac{k_2}{|w'_1|}) - \mathcal{C}(w'_1)(\frac{k_1+1}{|w'_1|}) = D - A$.

In any case, the conclusion follows the choice of $t'_1 = \frac{k_1+1}{|w'_1 w'_2|}$ and the corresponding $t'_2$ ($\frac{k_2}{|w'_1 w'_2|}$ or $\frac{k_2+1}{|w'_1 w'_2|}$ depending on the case).

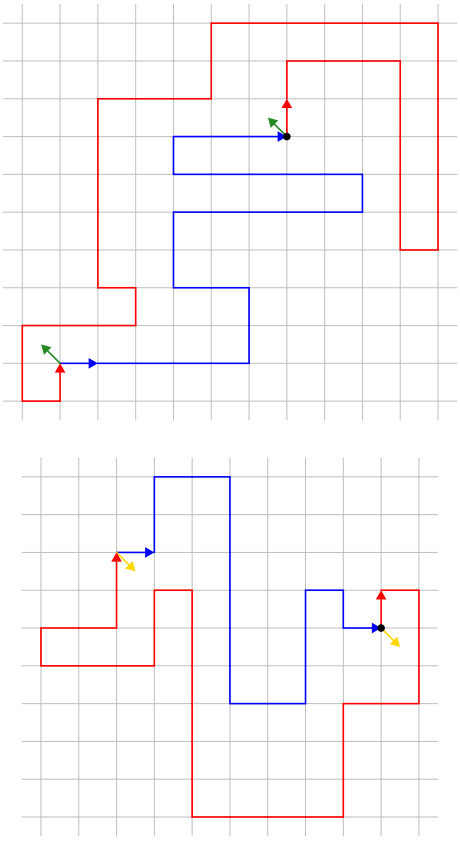


Figure 4: Examples of case 5

□

### *3.1. Well-nestedness*

A question that immediately arises from the proof we give that $MIX$ and $O_2$ are 2-MCFLs is whether the intricacy of case 5 is necessary for proving the

inria-00564552, version 1 - 9 Feb 2011

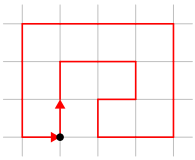


Figure 5: Curve representation of $aabb\bar{a}\bar{b}\bar{a}bbaaab\bar{b}\bar{b}\bar{b}\bar{a}\bar{a}\bar{a}b$

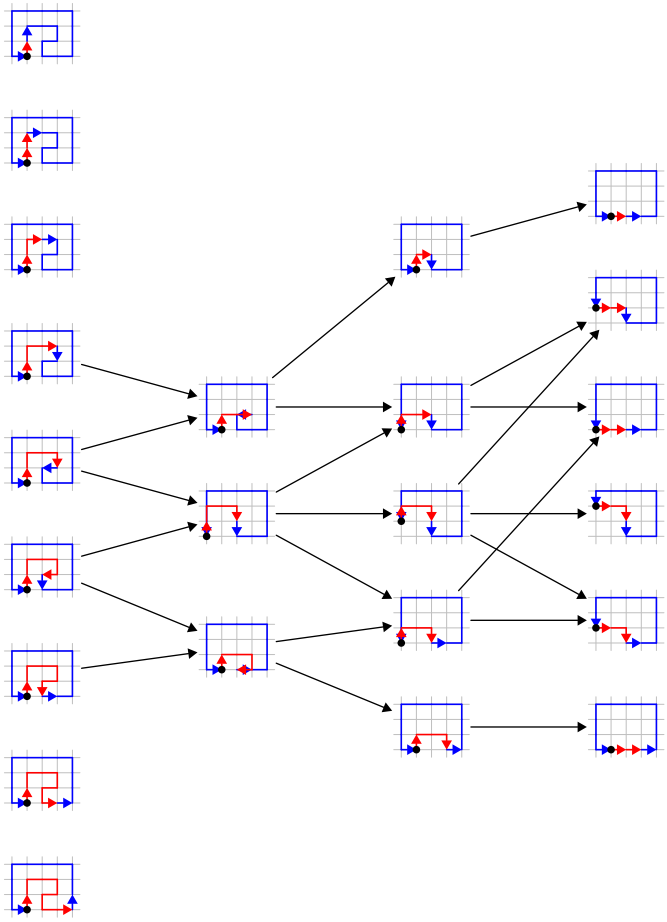


Figure 6: Parsing of $aabb\bar{a}\bar{b}\bar{a}bbaaab\bar{b}\bar{b}\bar{b}\bar{a}\bar{a}\bar{a}b$

result. A thing we can remark is that the four first cases only use rules that are well-nested to make the induction work. Indeed, case 1 is based on the rule:

$$\mathrm{Inv}(x_1x_2, y_1y_2) \text{ :- } \mathrm{Inv}(x_1, x_2),\ \mathrm{Inv}(y_1, y_2)$$

inria-00564552, version 1 - 9 Feb 2011

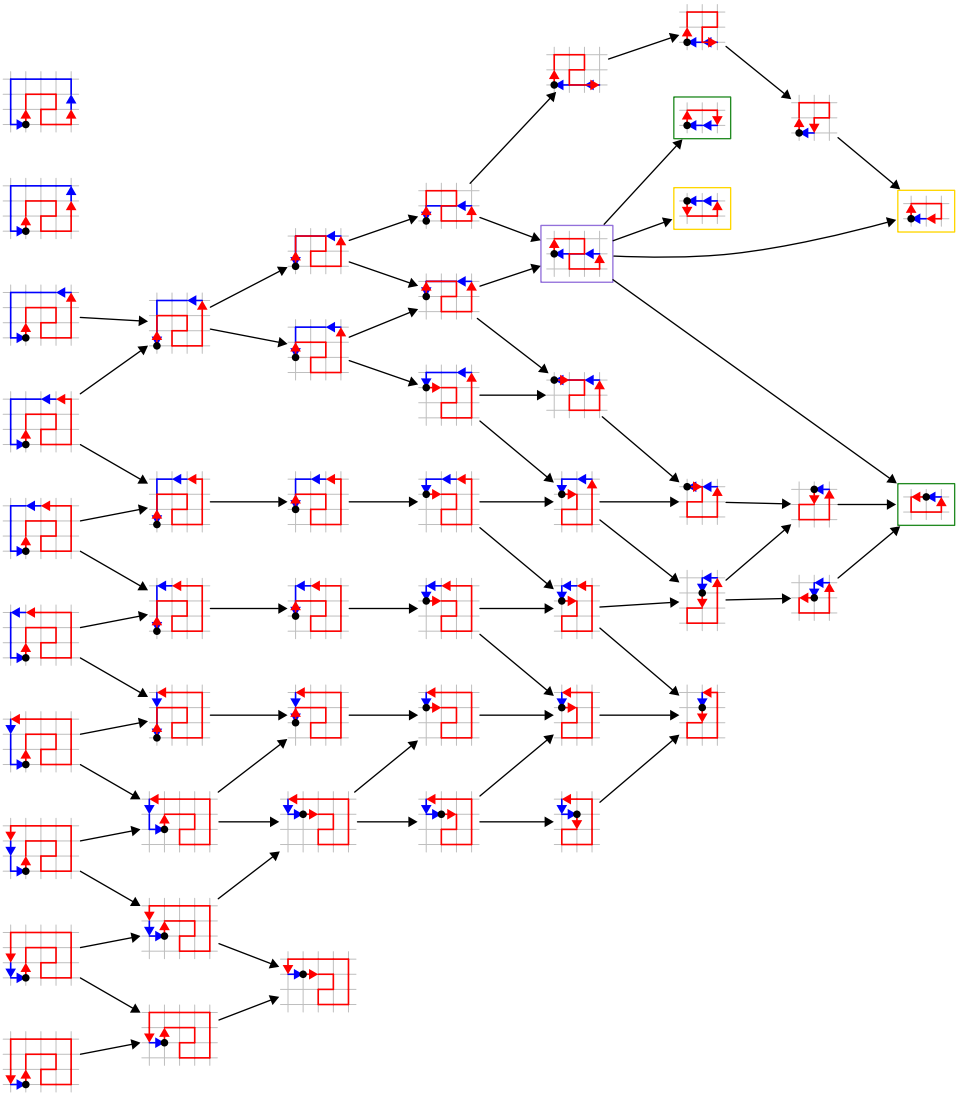


Figure 7: Parsing of $aabb\bar{a}\bar{b}\bar{a}bbaaa\bar{b}\bar{b}\bar{b}\bar{b}\bar{a}\bar{a}\bar{a}b$

while case 2 is based on the rules:

$$\begin{array}{lcl} \mathrm{Inv}(\alpha x_1 \bar{\alpha}, x_2) & :- & \mathrm{Inv}(x_1, x_2) \\ \mathrm{Inv}(x_1 \alpha, \bar{\alpha} x_2) & :- & \mathrm{Inv}(x_1, x_2) \\ \mathrm{Inv}(x_1, \alpha x_2 \bar{\alpha}) & :- & \mathrm{Inv}(x_1, x_2) \end{array}$$

inria-00564552, version 1 - 9 Feb 2011

where $\alpha$ is in $\{a; \bar{a}; b; \bar{b}\}$ and $\bar{\alpha} = \bar{a}$, $\bar{\alpha} = a$, $\bar{\alpha} = \bar{b}$ and $\bar{\alpha} = b$ when $\alpha = a$, $\alpha = \bar{a}$, $\alpha = b$ and $\alpha = \bar{b}$ respectively. Case 3 only requires the rules:

$$\begin{array}{lcl} \mathrm{Inv}(x_1x_2y_1, y_2) & :- & \mathrm{Inv}(x_1, x_2),\ \mathrm{Inv}(y_1, y_2) \\ \mathrm{Inv}(y_1x_1x_2, y_2) & :- & \mathrm{Inv}(x_1, x_2),\ \mathrm{Inv}(y_1, y_2) \\ \mathrm{Inv}(y_1, x_1x_2y_2) & :- & \mathrm{Inv}(x_1, x_2),\ \mathrm{Inv}(y_1, y_2) \\ \mathrm{Inv}(y_1, y_2x_1x_2) & :- & \mathrm{Inv}(x_1, x_2),\ \mathrm{Inv}(y_1, y_2) \end{array}$$

and that parts of case 4 that are disjoint from case 2 and case 3 only rely on the rules:

$$\mathrm{Inv}(x_1y_1, y_2x_2) :- \mathrm{Inv}(x_1, x_2),\ \mathrm{Inv}(y_1, y_2)$$

Whereas case 5 uses the rules:

$$\begin{array}{lcl} \mathrm{Inv}(x_1y_1x_2, y_2) & :- & \mathrm{Inv}(x_1, x_2),\ \mathrm{Inv}(y_1, y_2) \\ \mathrm{Inv}(x_1, y_1x_2y_2) & :- & \mathrm{Inv}(x_1, x_2),\ \mathrm{Inv}(y_1, y_2) \end{array}$$

The four first cases only rely on well-nested rules while the fifth one relies on non-well-nested rules. We here show an example of word, that is in $O_2$ but that is not in the language defined by the grammar where the non-well-nested rules have been omitted. This word is:

$$w = aabb\bar{a}\bar{b}\bar{a}bbaaa\bar{b}\bar{b}\bar{b}\bar{a}\bar{a}\bar{a}b$$

which is represented by the curve on figure 5. Figure 6 and figure 7 show all the possible ways of one can parse the word $w$ using the grammar where the non-well-nested rules have been removed. We have proceeded in splitting $w$ in all the possible ways into two strings $w_1$ and $w_2$ and we have drawn $w_1$ in red and $w_2$ in blue. Then we have recursively put arrows between possible premises that would allow to derive these pairs of strings. The only cases that show up in the course of this process are cases 2 and 4. The vast majority of premises are obtained using case 2 with the rules:

$$\begin{array}{lcl} \mathrm{Inv}(\alpha x_1\bar{\alpha}, x_2) & :- & \mathrm{Inv}(x_1, x_2) \\ \mathrm{Inv}(x_1\alpha, \bar{\alpha}x_2) & :- & \mathrm{Inv}(x_1, x_2) \\ \mathrm{Inv}(x_1, \alpha x_2\bar{\alpha}) & :- & \mathrm{Inv}(x_1, x_2) \end{array}$$

There are cases where both case 2 and case 4 can apply, but, in those cases, without loss of generality we may simply apply case 2. The only case where only case 4 applies, without possibly using rules of case 2, is that of the pair:

$$(abb\bar{a}\bar{b}\bar{a}bba, \bar{b}\bar{b}\bar{b})$$

which is surrounded with a purple line in figure 7. Its premises may be the pairs $(abb\bar{a}\bar{b}, \bar{b})$ and $(\bar{a}bba, \bar{b}\bar{b})$ surrounded with a yellow line, and the pairs $(abb\bar{a}, \bar{b}\bar{b})$ and $(\bar{b}\bar{a}bba, \bar{b})$, surrounded with a green line. For all the splitting of $w$ in $w_1$ and $w_2$, using the well-nested rules, we cannot obtain a derivation of $\mathrm{Inv}(w_1, w_2)$. In each case a derivation cannot be completed because it involves a pair of strings for which only case 5 can apply. This shows that the non-well-nested rules are mandatory in order to make the grammar capture $O_2$. We believe that this situation can be generalized so as to prove that there is no well-nested 2-MCFL (*i.e.* no Tree Adjoining Language) that can define $O_2$. Hopefully yet another generalisation would allow to prove that $O_2$ is not a well-nested MCFL.

inria-00564552, version 1 - 9 Feb 2011

## 4. A Theorem on Jordan curves

In this section we give the proof of Theorem 1. This Theorem is actually stronger than necessary to obtain Theorem 2. We could have taken advantage from the fact that the curves we are interested in are drawn on the grid. But we have preferred to prove a theorem that presents an interest on its own. The drawback of this generality is that it makes certain parts of the proof get slightly more complicated.

For proving Theorem 1, we shall assume, without loss of generality, that $A' = 0$ and $D' = 1$, and $D - A = 1$. The idea behind the proof is to use the complex exponential function (*i.e.* $\exp : z \to e^{2i\pi z}$) so as to observe certain properties of curves and arcs in the plane. A first simple property (Theorem 5) is that if a Jordan curve $f$ has $k$ elements of $\mathbb{Z}$ in its interior then its image through the exponential is winding $k$ times or $-k$ times (depending on the orientation of $f$ with respect to the chosen orientation of the plane) around 1 (note that for every $k$ in $\mathbb{Z}$, $\exp(k) = 1$). This implies (Corollary 4) in particular that the image by exp of the curve we consider in Theorem 1 is winding at least 2 times or at most $-2$ times around 1. A second property (Theorem 4) is that for a given Jordan arc $f$ such that $f(0) - f(1) = 1$ (*resp.* $f(0) - f(1) = -1$) the following properties are equivalent:

1. $f$ does not contain any strict (*i.e.* different from $f$) subarc $g$ such that $g(0) - g(1) = 1$ (*resp.* $g(0) - g(1) = -1$)
2. the image of $f$ by exp is a Jordan curve which is winding 1 (*resp.* $-1$) time around 0.

As a consequence a Jordan arc $f$ such that $f(0) - f(1) = 1$ (*resp.* $f(0) - f(1) = -1$) which does not contain any strict subarc $g$ such that $g(0) - g(1) = 1$ (*resp.* $g(0) - g(1) = -1$) is transformed by exp into a Jordan curve that winds 0 or 1 (*resp.* 0 or $-1$) time around 1. Thus, this second property implies that a Jordan curve that does not satisfy the conclusion of Theorem 1 is a Jordan curve that is composed by an arc $f_1$ and an arc $f_2$ such that their respective images by exp are winding 0 or 1 time and 0 or $-1$ around 1 which implies that the image of $f$ is winding between $-1$ and 1 time around 1. So that $f$ is contradicting the conclusion of Corollary 4 and therefore cannot satisfy the hypotheses of Theorem 1. From these facts Theorem 1 follows.

Most of the effort of the proof is dedicated to establishing the second property (Theorem 4). The difficulty of the proof is coming more particularly from Lemma 16 which shows the rather intuitive property that a closed curve $f$ winding strictly more than once around a point $p$ contains a subcurve that winds precisely once around $p$. The proof of Lemma 16 relies on a simple property of Jordan curves (Theorem 6), namely that a Jordan curve $f$ winds 1 time around all the points in its interior or $-1$ time. The structure of the proof of this Lemma is rather simple: it consists in removing the trivial subcurves of $f$ using an homotopy that only affects those subcurves, then we remove the subcurves of $f$ that are winding negatively around $p$. In doing so, we obtain a curve $g$ that winds more than $f$ around $p$, so that, by Lemma 6, it cannot be a Jordan curve,

inria-00564552, version 1 - 9 Feb 2011

furthermore, by construction, all subcurves of $g$ wind strictly positively around $p$. There is a natural partial order of inclusion on the subcurves of $g$ and we take a subcurve of $g$ that is minimal for this inclusion relation. As this subcurve is minimal, it must be a Jordan curve and as it is winding positively around $p$, it must, by Lemma 6, wind only once around it. Then it can be easily checked that this subcurve of $g$ induces a subcurve of $f$ that winds exactly once around $p$. Most of the technicalities we develop in the course of the proof of Theorem 1 is meant to give a precise meaning to the *removal* of trivial subcurves or subcurves winding negatively around $p$. Indeed these removals can be performed in many ways and in certain cases there is an uncountable number of ways of doing it and the number of trivial subcurves we may need to remove can be infinite.

The formalisation of these removals is based on the representation of subcurves as families of closed intervals included in $[0;1]$. We concentrate on their crucial property of being what we call *transitive systems of closed intervals* (Definition 2). We prove Lemma 12 which shows that we may extract from these families a subfamily that allows us to perform the removal. Because we deal with families that are *a priori* infinite the proof Lemma 12 heavily uses Zorn Lemma. Once we have obtained a family of trivial subcurves whose removal removes all the trivial subcurves, we homotop each of them to a constant curve (*i.e.* to the curve that is constantly equal to their base point). Thus the curve we obtain is piecewise stationary and in order to complete the removal we need to make it non-stationary. Here, a technicality, that mostly comes from the generality in which we wish to prove Theorem 1, shows up. It comes from the fact that the family of intervals on which the curve is stationary can be infinite. A quite pathological case can be that the cumulated measure (in Lebesgues sense) of the intervals composing this family is 1 (*i.e.* the measure of $[0;1]$!) meaning that the curve is stationary almost everywhere. As the situation is quite similar to Cantor function (see [GO90]), we use a construction (see Lemma 14) similar to the one of Cantor function in order to make the curve non-stationary. This overall construction has the nice property that every subarc of the curve we obtain is homotop to a subarc of the original curve (this is formalized through the notion of conservative deformation introduce in Definition 3). Then we just need to find a subcurve of this new curve that winds only once around $p$. As we previously explained, this is done by removing the subcurves that wind negatively around $p$ and taking a minimal subcurve of that new curve that winds positively around $p$. This minimal subcurve must be a Jordan curve and therefore must wind only once around $p$. To understand why this subcruve induces a subcurve winding once in the original curve, it suffices to remark that the only way it would not would be that, in the original curve, it was containing a subcurve winding negatively around $p$. But this would imply that in the original curve it would have been winding negatively around $p$ so that it should have been removed in the process of obtaining the new curve from the original one which leads to a contradiction.

The exposition of the proof is organized as follows. In the subsection 4.1 we present the topological spaces, the mapping between them and the basic properties we are going to use in the proof. In 4.2 we focus more particularly

inria-00564552, version 1 - 9 Feb 2011

on the punctured plane. We explain the relationship between its fundamental group and winding numbers and we give some basic properties about homotopy classes in term of winding numbers. Subsection 4.3 gives the tools that are necessary to proceed to the removal in a curve of its trivial subcurves or of its subcurves that wind negatively around a point. Because we represent subcurves as intervals this subsection mainly deals with properties of families intervals. The subsection 4.4 contains the proof of Lemma 16. And finally, 4.5 gives the proof of Theorem 1.

*4.1. Spaces, continuous functions and homomorphisms*

We present here the spaces which we are going to use in the proof of Theorem 1. These spaces and their fundamental groups are:

- the plane (we confuse the plane with the field of complex numbers) $\mathbb{C}$ whose fundamental group is trivial,
- the punctured plane $\mathbb{C}_p$ at $p$: $\mathbb{C} - \{p\}$ whose fundamental group is isomorphic to $\mathbb{Z}$,
- the twice punctured plane $\mathbb{C}_{p_1,p_2}$ at $p_1$, $p_2$: $\mathbb{C} - \{p_1; p_2\}$ whose fundamental group is isomorphic to the free group generated by two elements,
- the discretely punctured plane $\mathbb{C}_\mathbb{Z}$: $\mathbb{C} - \mathbb{Z}$ whose fundamental group is isomorphic to the free group generated by $\mathbb{Z}$.

For any point $z_1$, $z_2$ and $z_3$ respectively in $\mathbb{C}_p$, $\mathbb{C}_{p_1,p_2}$ and $\mathbb{C}_\mathbb{Z}$, we shall identify $\pi(z_0, \mathbb{C}_p)$ with $\mathbb{Z}$, $\pi(z_1, \mathbb{C}_{p_1,p_2})$ with the free group generated by $\beta_1$ and $\beta_2$, and $\pi(z_2, \mathbb{C}_\mathbb{Z})$ with the free group generated by the set $\{\alpha_k \mid k \in \mathbb{Z}\}$. Figure 8 shows the closed curves we choose to be the representatives of the homotopy classes corresponding to the generators of the fundamental groups of the topological spaces under consideration.

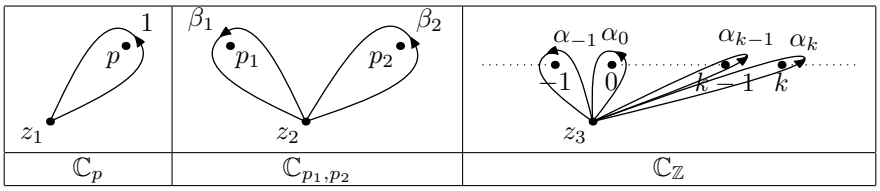



Figure 8: Representative curves of the classes denoted by the fundamental groups generators

With these conventions, we have the following properties:

1. Let $\iota_k : \mathbb{C}_\mathbb{Z} \to \mathbb{C}_k$ (with $k \in \mathbb{Z}$) be the inclusion mapping from $\mathbb{C}_\mathbb{Z}$ to $\mathbb{C}_k$. We have that $\hat{\iota}_k(\alpha_i) = \begin{cases} 1 \text{ when } i = k \\ 0 \text{ otherwise} \end{cases}$
2. Let $\rho_k : \mathbb{C}_{p_1,p_2} \to \mathbb{C}_{p_k}$, with $k$ in $\{1; 2\}$, be the inclusion mapping from $\mathbb{C}_{p_1,p_2}$ to $\mathbb{C}_{p_k}$. We have that $\hat{\rho}_k(\beta_j) = \begin{cases} 1 \text{ when } k = j \\ 0 \text{ otherwise} \end{cases}$

inria-00564552, version 1 - 9 Feb 2011

3. Let $\exp_p : \left\{ \begin{array}{ccc} \mathbb{C} & \rightarrow & \mathbb{C}_p \\ z & \rightarrow & e^{2i\pi z} + p \end{array} \right.$. When $f$ is a curve from $A$ to $B$ we have $B - A = k$ with $k \in \mathbb{Z}$ if and only if $exp_p \circ f$ is a closed curve and, $\widehat{exp_p}([f]_{\mathbb{C}}) = k$.
4. Let $\exp'_p : \left\{ \begin{array}{ccc} \mathbb{C}_{\mathbb{Z}} & \rightarrow & \mathbb{C}_{0,1} \\ z & \rightarrow & e^{2i\pi z} + p \end{array} \right.$. We have that $\widehat{\exp'_p}(\alpha_k) = \beta_1^j \beta_2 \beta_1^{-j}$, where $j$ is the integer part of the real part of $k - z_3$ ($z_3$ being the base point we have fixed for $\mathbb{C}_{\mathbb{Z}}$).

The properties numbered 1, 2 and 4 are obtained by computing the homotopy classes of the images obtained respectively with $\iota_k$, $\rho_k$ and $\exp'_p$ of the curves that we have chosen as representatives for the various spaces. For the property numbered 3, the *if part* is done similarly to the other properties; it suffices to compute the homotopy class of the image by $\exp_p$ of the line joining $A$ to $B$ which is obviously $k$. For the *only if* part, we need to notice that $\exp_p$ is covering projection (if suffices to remark that for every $z \in \mathbb{R}$, exp is an homeomorphism from $D_z = \{x + iy \mid z \leq x < z + 1 \wedge y \in \mathbb{R}\}$ to $\mathbb{C}_0$), and use the homotopy lifting property which allows reducing the problem to the computation of the homotopy classes of the circle passing through $z_1$ (the base point we fixed in $\mathbb{C}_p$) and centered in $p$, that are simply the lines from $z$ to $z + 1$ with $z$ in $\exp_p^{-1}(z_1)$ (notice that $z + 1$ is also in $\exp_p^{-1}(z_1)$).

*4.2. Fundamental group of $\mathbb{C}_p$, winding number and Jordan curves*

In the punctured plane $\mathbb{C}_p$, fixing an orientation of the plane[2], a closed curve $f$ with base point $A$ is associated a *winding number* $k$ when, according to the orientation, it winds $k$ times around $p$. A rigorous definition of the *winding number*, also called the *index*, is given in [AS60].

An easy Theorem (for a proof see [AS60]) that correlates the winding number and homotopy classes of the punctured plane is that two curves in the punctured plane are homotop to each other if and only if they have the same winding number.

**Theorem 3** *Given $f$ and $g$ two curves in the punctured plane $\mathbb{C}_p$ having the same base point, then the two following properties hold:*

1. *$[f]_{\mathbb{C}_p} = [g]_{\mathbb{C}_p}$ if and only if $f$ and $g$ have the same winding number,*
2. *the winding number of $f \cdot g$ is the sum of the winding numbers of $f$ and of $g$.*

This Theorem has the consequence that the homotopy classes of the punctured plane can be faithfully represented as the winding number of the curves. In the following we will often use the notion of winding numbers instead of the homotopy classes for closed curves in the punctured plane.

A Jordan curve splits the plane in two disconnected parts, one being bounded and the other being unbounded. The bounded part is called the *interior of the*

[2]In what follows we use the trigonometric or counter-clockwise orientation of the plane.

inria-00564552, version 1 - 9 Feb 2011

*curve* and the unbounded one is called the *exterior of the curve*. And the following Lemma is a consequence of the Theorem 10E of [AS60].

**Lemma 6** *Given a Jordan curve $f$ of $\mathbb{C}$, there is $k \in \{-1; 1\}$ such that for any $p$ of $\mathbb{C}$ we have:*

1. *if $p$ is in the interior of $f$ then $[f]_{\mathbb{C}_p} = k$*
2. *if $p$ is in the exterior of $f$ then $[f]_{\mathbb{C}_p} = 0$*

The Lemma says that a Jordan curve is winding one time around each points of its interior positively or negatively depending on its orientation and that also it is not winding (or winding zero time) around any points of its exterior.

**Lemma 7** *Given a closed curve $f$ in $\mathbb{C}_p$ then the winding number of $f$ is the same as the winding number of $f_{|[t;1]} \cdot f_{|[0;t]}$ for every $t$ in $[0; 1]$.*

The previous Lemma shows that the winding number of a closed curve in the punctured plane does not depend on the base point.

**Lemma 8** *Given a closed curve $f$ in $\mathbb{C}_p$, such that $f_{|[t_1;t_2]}$ is also closed then $[f]_{\mathbb{C}_p} = [f_{|[0;t_1]} \cdot f_{|[t_2;1]}]_{\mathbb{C}_p} + [f_{|[t_1;t_2]}]_{\mathbb{C}_p}$.*

*Proof.* From Lemma 7 we have that $f$ has the same winding number as $f_{|[t_1;1]} \cdot f_{|[0;t_1]}$ which is homotop to $f_{|[t_1;t_2]} \cdot f_{|[t_2;1]} \cdot f_{|[0;t_1]}$ and thus $[f]_{\mathbb{C}_p} = [f_{|[t_1;t_2]}]_{\mathbb{C}_p} + [f_{|[t_2;1]} \cdot f_{|[0;t_1]}]_{\mathbb{C}_p}$. But still using Lemma 7 we have obtain that $[f_{|[t_2;1]} \cdot f_{|[0;t_1]}]_{\mathbb{C}_p} = [f_{|[0;t_1]} \cdot f_{|[t_2;1]}]_{\mathbb{C}_p}$ and the result follows. □

**Lemma 9** *Given a closed curve $f$ in $\mathbb{C}_p$, and a finite family of intervals $\mathcal{C} = \{[t_{1,1}; t_{2,1}]; \ldots; [t_{1,n}; t_{2,n}]\}$ such that $t_{2,k} < t_{1,k+1}$ and $f_{|[t_{1,k};t_{2,k}]}$ is also a closed curve, then $[f]_{\mathbb{C}_p} = [f_{|[0;t_{1,1}]} \cdot \ldots f_{|[t_{2,k};t_{1,k+1}]} \cdot f_{|[t_{2,n};1]}]_{\mathbb{C}_p} + \sum_{k=1}^{n} [f_{|[t_{1,k};t_{2,k}]}]_{\mathbb{C}_p}$.*

*Proof.* This is obtained by a simple inductive use of Lemma 8. □

*4.3. On sets of closed subsets of* $[0; 1]$

This section is devoted to some rather technical matter which aims at studying the distribution of subcurves of a given curves. As subcurves are represented a restriction of arcs/curves to closed intervals of $[0; 1]$, this is the reason why it boils down to the study of properties of sets of closed subsets of $[0; 1]$. This part of our proof would not be necessary, in the context of only proving the Theorem 1 for curves of the form $\mathcal{C}(w)$ with $w$ in $O_2$.

**Lemma 10** *If $\mathcal{C}$ is a set of pairwise disjoint non-trivial closed intervals of $[0; 1]$ then, $\mathcal{C}$ is countable.*

*Proof.* Because the intervals $\mathcal{C}$ are pairwise disjoint and included in $[0; 1]$, there are finitely many of them with a length greater than $\frac{1}{n}$ for every $n$ in $\mathbb{N} - \{0\}$. Furthermore, because none of them is trivial, for each $C$ in $\mathcal{C}$, there is $n$ in $\mathbb{N}$ such that the length of $C$ is greater than $\frac{1}{n}$, from which we can infer that $\mathcal{C}$ is countable. □

inria-00564552, version 1 - 9 Feb 2011

**Definition 1.** Given a totally ordered set $K$, $(C_k)_{k\in K}$ is *an increasing (resp. decreasing) chain of closed intervals of* $[0;1]$, if, for $k_1$ and $k_2$ in $K$, $k_1 < k_2$ implies $C_{k_1} \subseteq C_{k_2}$ ($C_{k_2} \subseteq C_{k_1}$). The *limit* of $(C_k)_{k\in K}$ is the smallest closed interval of $[0;1]$ that contains $\bigcup_{k\in K} C_k$ (*resp.* is $\bigcap_{k\in K} C_k$).

**Lemma 11** *Given a totally ordered set $K$, $(C_k)_{k\in K}$ an increasing chain of closed intervals of $[0;1]$ and $J$ a subset of $[0;1]$ such that for every $k$ in $K$, $C_k \cap J = \emptyset$, if $C = [a;b]$ is the limit of $(C_k)_{k\in K}$ then $C_k \cap J$ is included in $\{a;b\}$.*

**Definition 2.** A *transitive system of closed intervals of* $[0;1]$, $\mathcal{C}$, is a set of non-trivial closed intervals of $[0;1]$ such that:

1. if $C_1 \cap C_2 = \{a\}$ for some $a$ in $[0;1]$, $C_1 \cup C_2$ is in $\mathcal{C}$,
2. for any increasing chain in $\mathcal{C}$, its limit is in $\mathcal{C}$.

**Lemma 12** *Given, $\mathcal{C}$, a transitive system of non-trivial closed intervals of $[0;1]$, there is a countable subset $\mathcal{C}'$ of $\mathcal{C}$ satisfying the following properties:*

1. *the elements of $\mathcal{C}'$ are pairwise disjoint,*
2. *for any $C$ in $\mathcal{C}$ there is $C'$ in $\mathcal{C}'$ such that $C \cap C' \neq \emptyset$, and if $C' \subseteq C$ then $C' = C$.*

*Proof.* Given $\mathcal{D}$ a subset of $\mathcal{C}$ we say that $\mathcal{D}$ verifies the property $\mathcal{P}$ (which we write $\mathcal{P}(\mathcal{D})$) if:

1. the elements of $\mathcal{D}$ are pairwise disjoint,
2. for $C$ in $\mathcal{C}$ either for every $D \in \mathcal{D}$, $C \cap D = \emptyset$, or there is $D$ in $\mathcal{D}$ such that $C \cap D \neq \emptyset$ and if $D \subseteq C$ then $D = C$.

Because, $\mathcal{C}$ is a transitive system of closed intervals, each of its increasing chain has an upper bound in $\mathcal{C}$. This has the consequence, by Zorn Lemma, that there is a $M$ in $\mathcal{C}$ such that $M$ is maximal for the inclusion in $\mathcal{C}$. Thus the set $\{M\}$ satisfies the property $\mathcal{P}$: the first condition is trivially fulfilled because $\{M\}$ is a singleton set and the second condition is a consequence of the maximality of $M$. This shows that there are subsets of $\mathcal{C}$ satisfying $\mathcal{P}$.

Let $G$ be a totally ordered set, and $(\mathcal{D}_g)_{g\in G}$ be an increasing chain of subsets of $\mathcal{C}$ that satisfy $\mathcal{P}$. We are going to see that $\mathcal{D} = \bigcup_{g\in G} \mathcal{D}_g$ is a subset of $\mathcal{C}$ that satisfies $\mathcal{P}$. Let's suppose that there are $D_1$ and $D_2$ in $\mathcal{D}$ such that $D_1 \cap D_2 \neq \emptyset$, then let $g_1$ and $g_2$ in $G$ be such that $D_1 \in \mathcal{D}_{g_1}$ and $D_2$ is in $\mathcal{D}_{g_2}$. If $g$ is the maximum value in $G$ between $g_1$ and $g_2$, then, because the chain $(\mathcal{D}_g)_{g\in G}$ is increasing, we have that both $D_1$ and $D_2$ are in $\mathcal{D}_g$, and as $\mathcal{P}(\mathcal{D}_g)$ holds, we have that $D_1 = D_2$. Let $C$ in $\mathcal{C}$ be such that there is $D$ in $\mathcal{D}$ such that $C \cap D \neq \emptyset$. We let $g$ in $G$ be such that $D \in \mathcal{D}_g$, since $\mathcal{P}(\mathcal{D}_g)$ and $C \cap D \neq \emptyset$, there is $D'$ in $\mathcal{D}_g$ such that $D' \cap C \neq \emptyset$ and if $D' \subseteq C$ then $D' = C$. As $D'$ is in $\mathcal{D}_g$, and thus in $\mathcal{D}$, this completes the proof that $\mathcal{P}(\mathcal{D})$ holds. Thus any increasing chains

inria-00564552, version 1 - 9 Feb 2011

of subsets of $\mathcal{C}$ satisfying $\mathcal{P}$ has an upper bound that satisfies $\mathcal{P}$. Using Zorn Lemma, we let $\mathcal{C}'$ be a maximal subset of $\mathcal{C}$ satisfying $\mathcal{P}$.

We are now going to prove that the family $\mathcal{C}'$ satisfies the conditions required by the Lemma. The first condition is verified simply because, $\mathcal{C}$ verifies $\mathcal{P}$. In order to verify the second condition, it suffices to show that for any $C$ in $\mathcal{C}$, there is $C'$ in $\mathcal{C}'$ such that $C \cap C' \neq \emptyset$. We proceed by contradiction, let's assume that there is $C$ in $\mathcal{C}$ such that for any $C'$ in $\mathcal{C}'$, $C \cap C' = \emptyset$. We say that an element $D$ of $\mathcal{C}$ has property $\mathcal{Q}$ when:

1. $C \subseteq D$,
2. for every $C'$ in $\mathcal{C}'$, $D \cap C' = \emptyset$

In case any increasing chain of intervals of $\mathcal{C}$ verifying $\mathcal{Q}$ has an upper bound, Zorn Lemma gives an interval $D$ that is maximum for the inclusion and that satisfies $\mathcal{Q}$. We let $\mathcal{D} = \mathcal{C}' \cup \{D\}$ and we verify that $\mathcal{P}(\mathcal{D})$ contradicting the maximality of $\mathcal{C}'$. Indeed, by construction the intervals of $\mathcal{D}$ are pairwise disjoint, and let $E$ in $\mathcal{C}$ be such that such that $E \cap \bigcup_{F \in \mathcal{D}} F \neq \emptyset$. In case there is $F$ is in $\mathcal{C}'$ such that $F \cap E \neq \emptyset$, because $\mathcal{C}'$ verifies $\mathcal{P}$, we have the existence of $F'$ in $\mathcal{C}'$ such that $E \cap F' \neq \emptyset$ and $E \subseteq F'$ implies $F' = E$. In case for each $C'$ in $\mathcal{C}'$, $E \cap C' = \emptyset$, as $E \cap \bigcup_{F \in \mathcal{D}} F \neq \emptyset$, we must have $D \cap E \neq \emptyset$ and if $D \subseteq E$, then $E$ verifies $\mathcal{Q}$ and the maximality of $D$ implies $E = D$.

In case there is an increasing chain of intervals of $\mathcal{C}$ verifying $\mathcal{Q}$ which does not have an upper bound satisfying $\mathcal{Q}$, we let $E = [a; b]$ be its limit (with $a \leq b$). Because $E$ does not satisfy $\mathcal{Q}$, we must have $E \cap \bigcup_{C' \in \mathcal{C}'} C' \neq \emptyset$ and by Lemma 11 $E \cap \bigcup_{C' \in \mathcal{C}'} C' \subseteq \{a; b\}$. Thus, given $C'$ in $\mathcal{C}'$, if $C' \cap E \neq \emptyset$, then we must have $C' = [c; a]$ with $0 \leq c \leq a$ or $C' = [b; c]$ with $b \leq c \leq 1$. We let $D$ be the union of $E$ with the intervals of $\mathcal{C}'$ that $E$ intersects; with the previous remark and the fact that $\mathcal{C}$ is a transitive system of closed interval, $D$ is in $\mathcal{C}$. By construction of $D$, if, for some $C'$ in $\mathcal{C}$, $D \cap C' \neq \emptyset$, then $C' \subseteq D$. But, as $C$ is included in $D$ and as $C$ cannot be included in $C'$, we have that for every $C'$ in $\mathcal{C}'$, $C' \cap D \neq \emptyset$ implies that $C' \subseteq D$ and $C' \neq D$. This contradicts the fact that $\mathcal{C}'$ satisfies $\mathcal{P}$.

In any case we have seen that the supposition that there is a $C$ in $\mathcal{C}$ such that $C \cap \bigcup_{C' \in \mathcal{C}'} C = \emptyset$ leads to a contradiction. This finally shows that $\mathcal{C}'$ is a subset of $\mathcal{C}$ satisfying the required conditions (the countability of $\mathcal{C}'$ is just a consequence of Lemma 10). □

**Lemma 13** *Given an arc $f$ in $\mathbb{C}_p$, we let:*

1. *$Z_f$ be the set of closed intervals $[t_1; t_2] \subseteq [0; 1]$ such that $t_1 < t_2$, $f(t_1) = f(t_2)$ and $[f_{|[t_1;t_2]}]_{\mathbb{C}_p} = 0$,*
2. *$N_f$ be the set of closed intervals $[t_1; t_2] \subseteq [0; 1]$ such that $f(t_1) = f(t_2)$, $[f_{|[t_1;t_2]}]_{\mathbb{C}_p} < 0$,*
3. *$P_f$ be the set of closed intervals $[t_1; t_2] \subseteq [0; 1]$ such that $f(t_1) = f(t_2)$, $[f_{|[t_1;t_2]}]_{\mathbb{C}_p} > 0$.*

*$Z_f$, $N_f$ and $P_f$ are transitive systems of closed intervals.*

inria-00564552, version 1 - 9 Feb 2011

*Proof.* We are only going to show that $Z_f$ is a transitive system of closed intervals,, the proofs concerning $N_f$ and $P_f$ are essentially the same.

By definition, $Z_f$ contains no trivial intervals.

If $[t_1;t_2]$ and $[t_2;t_3]$ with $t_1 < t_2 < t_3$ are in $Z_f$ then $[f_{|[t_1;t_3]}]_{\mathbb{C}_p} = [f_{|[t_1;t_2]}]_{\mathbb{C}_p} + [f_{|[t_2;t_3]}]_{\mathbb{C}_p} = 0$ which implies that $[t_1;t_3]$ is also in $Z_f$.

We finally show that the limit of any increasing chain of $Z_f$ is in $Z_f$. Given such a chain $([t_{1,k};t_{2,k}])_{k\in K}$ where $K$ is a totally ordered set and $t_{1;k} < t_{2,k}$. Thus, the sequences $(t_{1,k})_{k\in K}$ and $(t_{2,k})_{k\in K}$ are respectively decreasing and bounded below by 0, and increasing and bounded above by 1, we define $t_1$ as $\inf((t_{1,k})_{k\in K})$ and $t_2$ as $\sup((t_{2,k})_{k\in K})$ so that the limit of $([t_{1,k};t_{2,k}])_{k\in K}$ is $[t_1;t_2]$. We have seen (section 4.1) that $\exp_p : \mathbb{C} \to \mathbb{C}_p$ that we define as $\exp_p(z) = \exp(z) + p$ is a covering projection so that for every arc $g$ in $\mathcal{C}$, $[\exp_p \circ g]_{\mathbb{C}_p} = 0$ iff $g(0) = g(1)$. We let $g$ be a path of $\mathbb{C}$ such that $f = \exp_p \circ g$. Given $k$ in $K$, since $[f_{|[t_{1,k};t_{2,k}]}]_{\mathbb{C}_p} = 0$, we have that $g(t_{1,k}) = g(t_{2,k})$. So if we let $\varphi(x,y) = g(x) - g(y)$, $\varphi$ is a continuous function from $[0;1]^2$ to $\mathbb{C}$ and as for all $k$ we have $\varphi(t_{1,k}, t_{2,k}) = 0$, the continuity of $\varphi$ implies that $\varphi(t_1, t_2) = 0$. Therefore $g(t_1) = g(t_2)$ and $[f_{|[t_1;t_2]}]_{\mathbb{C}_p} = 0$ which implies that $[t_1;t_2]$ is in $Z_f$. □

**Lemma 14** *Given a set of pairwise disjoint non-trivial closed intervals of* $[0;1]$*,* $\mathcal{C}$*, such that* $\bigcup_{C\in\mathcal{C}} C \neq [0;1]$*, there is a continuous and increasing function,* $\varphi$*, from* $[0;1]$ *to* $[0;1]$ *such that:*

1. $\varphi(0) = 0$ *and* $\varphi(1) = 1$*,*
2. $\varphi$ *is constant on each* $C$ *in* $\mathcal{C}$
3. *if* $\varphi$ *is constant on an open interval* $U$ *then there is* $C$ *in* $\mathcal{C}$ *such that* $U \subseteq C$*.*

*Proof.* From Lemma 10, $\mathcal{C}$ is countable. The case where $\mathcal{C}$ is finite is easy, it suffices to define $\varphi$ as a piecewise linear function that is constant on each interval $C_i$. In that case the fact that $\bigcup_{C\in\mathcal{C}} C \neq [0;1]$ allows satisfying the constraint that $\varphi(0) = 0$ and $\varphi(1) = 1$, because $\bigcup_{C\in\mathcal{C}} C \neq [0;1]$.

So in case $\mathcal{C}$ is infinite, because it is countable, we can index its elements with natural numbers and obtain a family of non-trivial closed intervals $(C_n)_{n\in\mathbb{N}}$ such that $C$ in $\mathcal{C}$ if and only if there is a unique $n$ in $\mathbb{N}$ verifying $C = C_n$. Our construction is similar to the one of the Cantor function (see [GO90]). We shall define a family of open intervals $(I_w)_{w\in L}$ where $L$ is a subset of $\{0;1\}^*$. Let $k_1$ and $k_2$ be such that $C_{k_1} = [0;a]$ for some $a$ in $[0;1]$ and $C_{k_2} = [b;1]$ and let $F_1 = C_{k_1}$ and $F_2 = C_{k_2}$. In case there is no such $k_1$ we let $F_1 = [0;0]$ and similarly if there is no such $k_2$ we let $F_2 = [1;1]$. We then let $I_\epsilon = [0;1] - (F_1 \cup F_2)$, notice that $I_\epsilon$ cannot be empty because, $\bigcup_{j\in J} C_j \neq [0;1]$. If $I_w$ is defined, let $k_w$ be the smallest element of $\mathbb{N}$ such that $C_{k_w}$ is included in $I_w$, if such a $k_w$ does not exist then for all $v \in \{0;1\}^+$, $wv$ is not in $L$; but if such a $k_w$ exists then $w0$ and $w1$ are in $L$ and if $I_w = (a_{I_w}; b_{I_w})$ and $C_{k_w} = [a_{k_w}; b_{k_w}]$, then $I_{w0} = (a_{I_w}; a_{k_w})$ and $I_{w1} = (b_{k_w}; b_{I_w})$. Note that $I_{w0} \cup I_{w1} = I_w - C_k$, that

inria-00564552, version 1 - 9 Feb 2011

whenever $w$ is in $L$, all the left factors of $w$ are in $L$ and that $w0$ is in $L$ if and only if is $w1$ is in $L$. A simple induction gives that if $w$ and $wv$ are in $L$ then $I_{wv} \subseteq I_w$ and that the inclusion is strict when $v$ is not the empty string. Similarly, we obtain that if $w_1$ and $w_2$ are both in $L$ and that neither $w_1$ is a left factor of $w_2$ nor $w_2$ is a left factor of $w_1$ then $I_{w_1} \cap I_{w_2} = \emptyset$. Finally note that for each $n$ in $\mathbb{N}$ (provided $n \neq k_1$ and $n \neq k_2$ in case either $k_1$ or $k_2$ exist), there is $w$ such that $k_w = n$.

We are now going to define a sequence of sets $(U_n)_{n\in\mathbb{N}}$ such that $\bigcap_{n\in\mathbb{N}} U_n = [0;1] - \bigcup_{n\in\mathbb{N}} C_n$. Given $n$ in $\mathbb{N}$, we define $L_n$ to be the set of strings in $L$ which have length $n$. We then inductively $(U_n)_{n\in\mathbb{N}}$ by $U_0 = [0;1]$ and $U_{n+1} = U_n - \bigcup_{w\in L_n} C_{k_w}$. Thus $U_{n+1} \subseteq U_n$ and it is quite easy to prove that $\bigcap_{n\in\mathbb{N}} U_n = [0;1] - \bigcup_{n\in\mathbb{N}} C_n$.

Finally we define $(\varphi_n)_{n\in\mathbb{N}}$ to be a sequence of piecewise linear increasing and continuous functions from $[0;1]$ to $[0;1]$. More precisely $\varphi_n$ is linear and strictly increasing on each connected component of $U_n$ (*i.e.* on each $I_w$ such that $w \in L_n$) and constant on $[0;1] - U_n$. We let $\varphi_0$ be the piecewise linear function that is constant outside $U_0$ and such that $\varphi_0(0) = 0$ and $\varphi_1(1) = 1$. The function $\varphi_{n+1}$ is the function that is equal to $\varphi_n$ on $[0;1] - \bigcup_{w\in L'_n} I_w$ and for every $w$ in $L'_n$ we define the restriction of $\varphi_{n+1}$ to $[a_{I_w}; b_{I_w}]$ as follows (see figure 9):

1. on $[a_{I_w}; a_{k_w}]$, $\varphi_{n+1}$ is the linear function such that $\varphi_{n+1}(a_{I_w}) = \varphi_n(a_{I_w})$ and $\varphi_{n+1}(a_{k_w}) = \frac{\varphi_n(a_{I_w})+\varphi_n(b_{I_w})}{2}$
2. on $[a_{k_w}; b_{k_w}]$, $\varphi_{n+1}$ is the function constantly equal to $\frac{\varphi_n(a_{I_w})+\varphi_n(b_{I_w})}{2}$,
3. on $[b_{k_w}; b_{I_w}]$, $\varphi_{n+1}$ is the linear function such that $\varphi_{n+1}(b_{I_w}) = \varphi_n(b_{I_w})$ and $\varphi_{n+1}(b_{k_w}) = \frac{\varphi_n(a_{I_w})+\varphi_n(b_{I_w})}{2}$

A simple induction on $n$ shows that $\varphi_n$ has the expected properties and that for each $w \in L$ such that $|w| = n$, then $\varphi_n(b_{I_w}) - \varphi_n(a_{I_w}) = 2^{-n}$ which implies that $\sup_{t\in[0;1]}(|\varphi_{n+1}(t) - \varphi_n(t)|) \leq 2^{-n-1}$ (see figure 9).

Thus the sequence of continuous functions $(\varphi_n)_{n\in\mathbb{N}}$ is uniformly convergent and we let $\varphi$ be the continuous function that is the limit of this sequence. Since for all $n$ in $\mathbb{N}$, $\varphi_n$ is an increasing function from $[0;1]$ to $[0;1]$ such that $\varphi_n(0) = 0$ and $\varphi_n(1) = 1$ and that is constant exactly on the interval $C_{k_w}$ with $w$ in $L_n$, and since $\bigcup_{n\in\mathbb{N}} = \bigcup_{w\in L_n} C_{k_w} = \bigcup_{n\in\mathbb{N}} C_n$ we have that $\varphi$, is an increasing continuous function from $[0;1]$ to $[0;1]$ that is constant exactly on the intervals of the family $(C_n)_{n\in\mathbb{N}}$ and such that $\varphi(0) = 0$ and $\varphi(1) = 1$, which is precisely what we were expecting. □

### *4.4. Subcurves of curves in $\mathbb{C}_p$*

In this section, we prove the a technical Lemma that contains most of the difficulties of the proof of Theorem 1.

Before proving the Lemma, we need to prove a technical property that generalizes to arcs in the plane the Corollary 1. This generalization is not trivial at all, since the developments of section 4.3 have been made mostly on the purpose proving it. But before generalizing the result, we first need to extend the

inria-00564552, version 1 - 9 Feb 2011

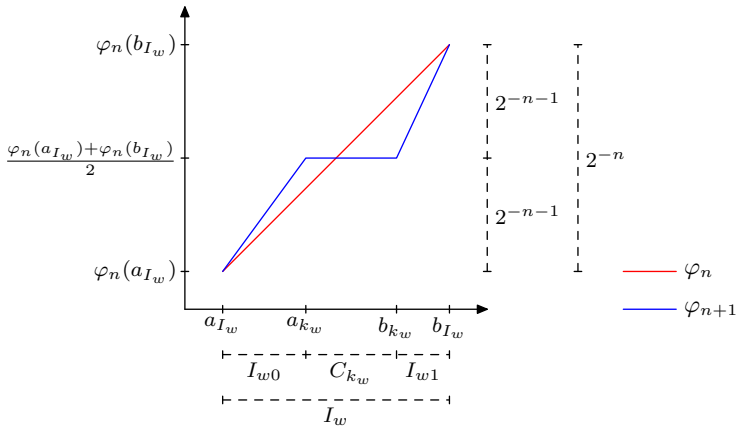



Figure 9: Given $w \in L'_n$, the graphs of $\varphi_n$ and $\varphi_{n+1}$ on $I_w$

notion of simplification for words to arcs by introducing the notion *conservative homotopy* or *conservative deformation*.

**Definition 3.** An homotopy $H$ that transforms $f$ into $g$ in a space $X$ is said *conservative* when for every $t_1$, $t_2$ in $[0;1]$ such that there is $t'_1$ and $t'_2$ in $[0;1]$ such that $g_{|[t_1;t_2]} \approx_X f_{|[t'_1;t'_2]}$. Then, $g$ is said to be a *conservative deformation of* $f$, which we write $g \prec_X f$

Obviously if $f_3 \prec_X f_2$ and $f_2 \prec_X f_1$ then $f_1$ then $f_3 \prec_X f_1$. The relation $\prec_X$ is a partial order inside homotopy classes.

We now prove the generalization of Corollary 1 for arcs.

**Lemma 15** *Given a non-trivial arc $f$, there is an arc $g$ such that $g \prec_X f$ and $g$ has no trivial subcurve.*

*Proof.* The proof is done in two steps. We first homotop $f$ so that we obtain an arc $h$ with the property that $h \prec_X f$ and whenever $h_{|[t_1;t_2]}$ is a trivial subcurve of $h$, then $h$ is constant on $[t_1;t_2]$. We then construct an arc $g$ from $h$ using a conservative homotopy that removes the interval where $h$ is stationary.

From Lemma 13 we have that $Z_f$ is a transitive system of closed intervals, Lemma 12 implies that there is a family $(C_j)_{j\in J}$ such that:

1. $J$ is countable and for every $j \in J$, $C_j \in Z_f$,
2. given $j_1$ and $j_2$ in $J$, $C_{j_1} \cap C_{j_2} \neq \emptyset$ iff $j_1 = j_2$,
3. for every $C$ in $Z_f$, there is $j$ in $J$ such that $C \cap C_j \neq \emptyset$ and $C_j \subseteq C$ implies $C_j = C$,
4. $\bigcup_{j\in J} C_j \neq [0;1]$ (because $f$ is non-trivial).

Given $j$ in $J$, since $[f_{|C_j}]_{\mathcal{C}_p} = 0$, $f_{|C_j}$ can be homotoped to a trivial curve $f'_j$ such that $f'_j(t) = f_{|C_j}(0)$ for all $t$ in $[0;1]$. We write $H_j(s)(t)$ an homotopy

inria-00564552, version 1 - 9 Feb 2011

transforming $f_{|C_j}$ into $f'_j$. As the family $(C_j)_{j\in J}$ is made of pairwise disjoint intervals, we may define $H$ as the following homotopy:

$$H(s)(t) = \begin{cases} H_j(s)(\frac{t-t_1}{t_2-t_1}) \text{ if } t \in C_j = [t_1;t_2] \text{ with } j \in J \\ f(t) \text{ otherwise} \end{cases}$$

We let $h(t) = H(1)(t)$. We first check that $h$ is a conservative deformation of $f$, *i.e.* for any $t_1$ and $t_2$ from $[0;1]$ such that $t_1 \leq t_2$, then there are $t'_1$ and $t'_2$ in $[0;1]$ such that $t'_1 \leq t'_2$ and $h_{|[t_1;t_2]} \approx_{\mathcal{C}_p} f_{|[t'_1;t'_2]}$. We let in case $t_1$ is in $C_{j_1}$ for some $j_1$ in $J$ we let $t'_1 = \inf\{t \mid t \in C_{j_1}\}$ and otherwise $t'_1 = t_1$. Similarly, in case $t_2$ is in $C_{j_2}$ for some $j_2$ in $J$ we let $t'_2 = \sup\{t \mid t \in C_{j_2}\}$ and otherwise $t'_2 = t_2$. Remark that in any cases, $[t_1;t_2] \subseteq [t'_1;t'_2]$, that if for some $j$ in $J$, we have $C_j \cap [t'_1;t'_2] \neq \emptyset$, we have $C_j \subseteq [t'_1;t'_2]$, and also that $h_{|[t'_1;t_1]}$ and $h_{|[t_2;t'_2]}$ are both trivial curves. This implies that $h_{|[t_1;t_2]} \approx_{\mathbb{C}_p} h_{|[t'_1;t_1]} \cdot h_{|[t_1;t_2]} \cdot h_{|[t_2;t'_2]} \approx_{\mathbb{C}_p} h_{|[t'_1;t'_2]}$ and as $H(s)_{|[t'_1;t'_2]}$ is an homotopy that transforms $f_{|[t'_1;t'_2]}$ into $h_{|[t'_1;t'_2]}$ we obtain that $h_{|[t_1;t_2]} \approx_{\mathbb{C}_p} f_{|[t'_1;t'_2]}$. If furthermore we have $[h_{|[t_1;t_2]}]_{\mathbb{C}_p} = 0$, then we must have $[f_{|[t'_1;t'_2]}]_{\mathbb{C}_p} = 0$ and therefore, there is $j$ in $J$ such that $C_j \cap [t'_1;t'_2] \neq \emptyset$. But we have seen that $C_j$ must be a subset of $[t'_1;t'_2]$ which implies that $[t'_1;t'_2] = C_j$ and $[t_1;t_2] \subseteq C_j$. Thus, $[h_{|[t_1;t_2]}]_{\mathbb{C}_p} = 0$ and $t_1 \neq t_2$ if and only if $[t_1;t_2] \subseteq \bigcup_{j\in J} C_j$.

We write $x \sim y$ when $h$ is constant on the interval $[x;y]$. The relation $\sim$ is an equivalence relation on $[0;1]$ and we write $\widetilde{x}$ for the equivalence class of $x$. From what we have just seen above, it is obvious that the classes of $\sim$ that are not singleton sets are precisely the intervals of the family $(C_j)_{j\in J}$.

Lemma 14 implies the existence of a continuous and increasing function, $\varphi$, from $[0;1]$ to $[0;1]$ such that:

1. $\varphi(0) = 0$ and $\varphi(1) = 1$,
2. $\varphi$ is constant on each $C_j$ for every $j$ in $J$
3. if $\varphi$ is constant on an open interval $U$ then there is $j$ in $J$ such that $U \subseteq C_j$.

Then the function $\psi$ from $[0;1]/\sim$ to $[0;1]$ such that $\psi(\tilde{t}) = \varphi(t)$ defines a homeomorphism between $[0;1]/\sim$ (equipped with the quotient topology) and $[0;1]$. The definition of $\psi$ implies that

$$\psi^{-1}(\varphi(t)) = \tilde{t} \tag{3}$$

If we let $h'$ be the continuous function from $[0;1]/\sim$ to $X$ such that $h'(\tilde{t}) = h(t)$, we then define $g$ to be $h' \circ \psi^{-1}$. The definition of $h$ and identity (3) imply

$$g(\varphi(t)) = h'(\psi^{-1}(\varphi(t))) = h(t) \tag{4}$$

We first show $g$ is a conservative deformation of $h$. For this we are going to prove that for each $t_1$ and $t_2$ in $[0;1]$ such that $t_1 \leq t_2$, there is $t'_1$ and $t'_2$ in $[0;1]$ with $t'_1 \leq t'_2$ such that $g_{[t_1;t_2]} \approx_{\mathbb{C}_p} h_{|[t'_1;t'_2]}$. Indeed, let $t'_1$ and $t'_2$ be such that $\varphi(t'_1) = t_1$ and $\varphi(t'_2) = t_2$ (such $t'_1$ and $t'_2$ exist since $\varphi$ is a surjection). Using (4) we have that $g(t_1) = g(\varphi(t'_1)) = h(t'_1)$ and $g(t_2) = g(\varphi(t'_2)) = h(t'_2)$, so let

$$G(s,t) = g(s((1-t)t_1 + tt_2) + (1-s)(\varphi((1-t)t'_1 + tt'_2)))$$

inria-00564552, version 1 - 9 Feb 2011

$G$ is obviously continuous and we have:

$$G(s,0) = g(st_1 + (1-s)\varphi(t'_1)) = g(st_1 + (1-s)t_1) = g(t_1) \tag{5}$$

and

$$G(s,1) = g(st_2 + (1-s)\varphi(t'_2)) = g(st_2 + (1-s)t_2) = g(t_2) \tag{6}$$

and, using (4) again,

$$G(0,t) = g(\varphi(tt'_1 + (1-t)t'_2)) = h(tt'_1 + (1-t)t'_2) = h_{|[t'_1;t'_2]} \tag{7}$$

and

$$G(1,t) = g(tt_1 + (1-t)t_2) = g_{|[t_1;t_2]} \tag{8}$$

The identities (5), (6), (7) and (8) show that $h_{|[t'_1;t'_2]} \approx_{\mathbb{C}_p} g_{|[t_1;t_2]}$ which finally implies that $g$ is a conservative deformation of $h$.

It remains to check that $g$ that if $g(t_1) = g(t_2)$ and $[g_{|[t_1;t_2]}]_{\mathbb{C}_p} = 0$, then $t_1 = t_2$. Given such $t_1$ and $t_2$, let $t'_1$ and $t'_2$ be such that $\varphi(t'_1) = t_1$ and $\varphi(t'_2) = t_2$. We have seen above that in such a case $g_{|[t_1;t_2]} \approx_{\mathbb{C}_p} h_{|[t'_1;t'_2]}$ which implies that $[h_{|[t'_1;t'_2]}]_{\mathbb{C}_p} = 0$ and that $h$ is constant on $[t'_1;t'_2]$ which implies that $t'_1 \sim t'_2$ and $\widetilde{t'_1} = \widetilde{t'_2}$. But, $t_1 = \varphi(t'_1)$ and $t_2 = \varphi(t'_2)$ so that, with (3), $\psi^{-1}(t_1) = \psi^{-1}(\varphi(t'_1)) = \widetilde{t'_1} = \widetilde{t'_2} = \psi^{-1}(\varphi(t'_2)) = \psi^{-1}(t_2)$ which finally gives $t_1 = t_2$ because $\psi$ is a homeomorphism. This implies that $g$ has no trivial subcurve. □

The following Lemma plays a key role in the proof of Theorem 1. It can be interpreted as a kind of intermediate value Theorem for arcs in the plane. Indeed, as it says that when a closed curve $f$ is doing $k$ turns, with $k > 1$, around around $p$ in the punctured plane $\mathbb{C}_p$, then there is a subcurve of $f$ that is doing exactly one turn around $p$ in $\mathbb{C}_p$. As the mapping $\varphi$ from $\mathbb{C}$ to $\mathbb{C}_p$ such that $\varphi(z) = e^{2i\pi z/v} + p$ is a covering projection, using the unique path-lifting property, we have that $f$ can be seen as the image of a unique (modulo the choice of $g(0)$ in $\varphi^{-1}(f(0))$) curve $g$ such that $g(1) - g(0) = kv$ . What is implied by the Lemma is that there is a subarc $h$ of $g$ such that $h(1) - h(0) = v$. However, it is not the case that for every natural number $l \leq k$ there is a subarc $h_l$ of $g$ such that $h_l(1) - h_l(0) = lv$. For a word $w$ in $\{a; \overline{a}; b; \overline{b}\}$, this Lemma, together with Fact 2 implies that if $\varphi(w) = kz$ with $k$ in $\mathbb{N} - \{0\}$, then there is a $u$, a factor of $w$, such that $\varphi(u) = z$.

**Lemma 16** *Given a closed curve $f$ in $\mathbb{C}_p$ whose winding number is $k$ such that $k > 1$ (resp. $k < -1$), then there is $t_1$, $t_2$ in $[0;1]$ such that, $t_1 < t_2$, $[t_1;t_2] \subsetneq [0;1]$, and $f_{|[t_1;t_2]}$ is a closed curve whose winding number is $1$ (resp. $-1$).*

*Proof.* We will only study the case where $k > 1$, the case where $k < -1$ being symmetric.

From Lemma 15 we have $g$ such that $g \prec_{\mathbb{C}_p} f$ and $g$ contains no trivial subcurves. Recall that $N_g$ is the set $\{[t_1;t_2] \mid [g_{|[t_1;t_2]}]_{|\mathbb{C}_p} < 0\}$; Lemmas 13 and 12 imply that there is a countable subset $N$ of $N_g$ which is a transitive system of closed intervals such that:

inria-00564552, version 1 - 9 Feb 2011

1. the elements of $N$ are pairwise disjoint,
2. for every $C$ in $N_g$ there is $D$ in $N$ such that, $C \cap D \neq \emptyset$ and $D \subseteq C$ implies $C = D$.

By uniform continuity of $g$, it is easy to see that there is $\rho$ such that every $[t_1; t_2]$ in $\mathcal{N}_g$ is such that $t_2 - t_1 > \rho$. This implies that $N$ is finite. If $N$ is equal to $\{[t_{1,1}; t_{2,1}]; \ldots; [t_{1,n}; t_{2,n}]\}$ with $t_{1,k} < t_{2,k} < t_{1,k+1}$ for $0 < k < n$, then using Lemma 9 we get that

$$[g]_{\mathbb{C}_p} = [g_{|[0;t_{1;1}]} \cdot \ldots \cdot g_{|[t_{2,n};1]}]_{\mathbb{C}_p} + \sum_{k=1}^{n} [g_{|[t_{1,k};t_{2,k}]}]_{\mathbb{C}_p}$$

If we let $g' = g_{|[0;t_{1;1}]} \cdot \ldots \cdot g_{|[t_{2,n};1]}$ since for every $k$, $[g_{|[t_{1,k};t_{2,k}]}]_{\mathbb{C}_p} < 0$, we get that the winding number of $[g']_{\mathbb{C}_p} \geq [g]_{\mathbb{C}_p} > 0$. Thus, from Lemma 6, the set $\mathcal{I}_{g'} = \{[t_1; t_2] \mid t_1 \neq t_2 \wedge g'(t_1) = g'(t_2)\}$ cannot be empty. But, from the construction of $g'$, we have that $Z_{g'} = \emptyset$ and that $N_{g'} = \emptyset$. Thus for every $[t_1; t_2]$ in $\mathcal{I}_{g'}$ we have that $[g'_{|[t_1;t_2]}]_{\mathbb{C}_p} > 0$. One can easily see that every decreasing chain (for the inclusion) in $\mathcal{I}_{g'}$ has a lower bound in $\mathcal{I}_{g'}$, thus, with Zorn Lemma, we can take a minimal element $[t_1; t_2]$ in $\mathcal{I}_{g'}$. From the minimality of $[t_1; t_2]$ we get that $g'_{|[t_1;t_2]}$ is a Jordan curve, and since $[g'_{|[t_1;t_2]}]_{\mathbb{C}_p} > 0$, Lemma 6 we must have $[g'_{|[t_1;t_2]}]_{\mathbb{C}_p} = 1$. But as $g' = g_{|[0;t_{1;1}]} \cdot \ldots \cdot g_{|[t_{2,n};1]}$ either $g'_{|[t_1;t_2]}$ is equal to $g_{|[u_1;u_2]}$ with $[u_1; u_2]$ included in $[t_{2,k-1}; t_{1,k}]$ for $k$ in $[n+1]$ and in such a case we have proved the Lemma, or there is $k_1$ and $k_2$ in $[n+1]$, such that $k_1 < k_2$ and for some $u_1$ in $[t_{2,k_1-1}; t_{1,k_1}]$ and some $u_2$ in $[t_{2,k_2-1}, t_{1,k_2}]$ we have $g'_{|[t_1;t_2]} \approx_{\mathbb{C}_p} g_{|[u_1;t_1,k_2]} \cdot \ldots \cdot g_{|[t_{2,t_2-1};u_2]}$ but we have that:

$$\begin{aligned} [g_{|[u_1;u_2]}]_{\mathbb{C}_p} &= [g'_{|[t_1;t_2]}]_{\mathbb{C}_p} + \sum_{[k_1 \leq k \leq k_2} [g_{|[t_{1,k};t_{2,k}]}]_{\mathbb{C}_p} \\ &= 1 + \sum_{k_1 \leq k \leq k_2} [g_{|[t_{1,k};t_{2,k}]}]_{\mathbb{C}_p} \\ &\leq 0 \end{aligned}$$

which contradicts the fact that $\mathcal{Z}_{g'} \cup \mathcal{N}_{g'}$ is empty. This then implies that $[g_{|[u_1;u_2]}]_{\mathbb{C}_p} = 1$. And since $g$ is a conservative deformation of $f$, the conclusion follows. □

A corollary of this Lemma is the following.

**Corollary 3.** *If $f$ is a closed curve in $\mathbb{C}_p$ whose winding number is 1 (resp. $-1$) and $f$ is not simple then there are $t_1, t_2 \in [0; 1]$ such that, $t_1 < t_2$, $[t_1; t_2] \subsetneq [0; 1]$ and $f_{|[t_1;t_2]}$ has winding number 0 or 1 (resp. $-1$).*

*Proof.* Let $t_1$ and $t_2$ be such that $t_1 < t_2$, $[t_1; t_2] \subsetneq [0; 1]$ and $f(t_1) = f(t_2)$ ($t_1$ and $t_2$ exist since $f$ is supposed not to be simple), then $f_{|[t_1;t_2]]}$ is a curve that has $k$ as winding number. If $k = 0$ then the conclusion is obvious. If $k > 0$ then we can conclude using the previous Lemma. If $k < 0$ then thanks

inria-00564552, version 1 - 9 Feb 2011

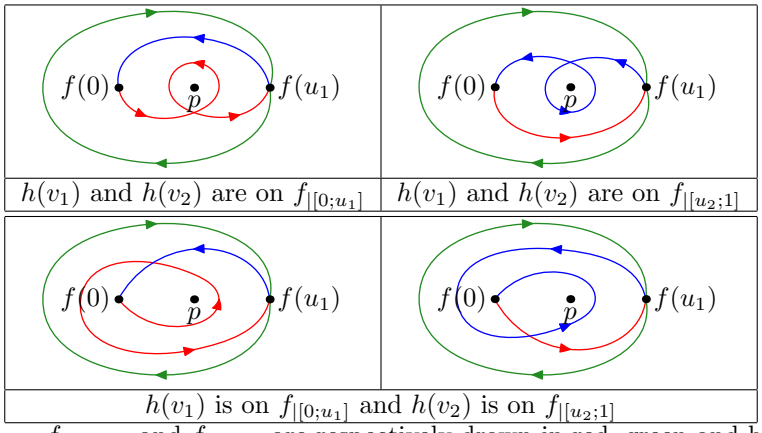
to the previous Lemma, we have the existence of $u_1$ and $u_2$ in $[t_1;t_2]$ such that $u_1 < u_2$ and $f_{|[u_1;u_2]}$ has winding number $-1$. Then we can see that $f = f_{|[0;u_1]} \cdot f_{|[u_1;u_2]} \cdot f_{|[u_2;1]}$. By hypothesis, we have that

$$\begin{aligned}[f]_{\mathbb{C}_p} &= [f_{|[0;u_1]} \cdot f_{|[u_1;u_2]} \cdot f_{|[u_2;1]}]_{\mathbb{C}_p} \\ &= [f_{|[0;u_1]}]_{\mathbb{C}_p} + [f_{|[u_1;u_2]}]_{\mathbb{C}_p} + [f_{|[u_2;1]}]_{\mathbb{C}_p} \\ &= 1\end{aligned}$$

but $h = f_{|[0;u_1]} \cdot f_{|[u_2;1]}$ is a closed curve and, since $[f_{|[u_1;u_2]}]_{\mathbb{C}_p} = -1$, we have $[h]_{\mathbb{C}_p} = 2$. This means, using the previous Lemma, that there is $v_1$, $v_2$ in $[0;1]$ such that $v_1 < v_2$ and $h_{|[v_1;v_2]}$ has winding number 1. There are three cases (*c.f.* figure 10):

1. $h \approx_{\mathbb{C}_p} f_{|[0;v_1']} \cdot f_{|[v_1';v_2']} \cdot f_{|[v_2';u_1]} \cdot f_{|[u_2;1]}$ with $h_{|[v_1;v_2]]} \approx_{\mathbb{C}_p} f_{|[v_1';v_2']}$ and the conclusion follows easily,
2. $h \approx_{\mathbb{C}_p} f_{|[0;u_1]} \cdot f_{|[u_2;v_1']} \cdot f_{|[v_1';v_2']} \cdot f_{|[v_2';1]}$ with $h_{|[v_1;v_2]]} \approx_{\mathbb{C}_p} f_{|[v_1';v_2']}$ and, here again the conclusion follows easily,
3. $h \approx_{\mathbb{C}_p} f_{|[0;v_1']} \cdot f_{|[v_1';u_1]} \cdot f_{|[u_2;v_2']} \cdot f_{|[v_2';1]}$ with $h_{|[v_1;v_2]]} \approx_{\mathbb{C}_p} f_{|[v_1';u_1]} \cdot f_{|[v_2';u_2']}$ but in that case we have

$$\begin{aligned}[f_{|[v_1';v_2']}]_{\mathbb{C}_p} &= [f_{|[v_1';u_1]} \cdot f_{|[u_1;u_2]} \cdot f_{|[u_2;v_2']}]_{\mathbb{C}_p} \\ &= [f_{|[v_1';u_1]} \cdot f_{|[u_2;v_2']}]_{\mathbb{C}_p} + [f_{|[u_1;u_2]}]_{\mathbb{C}_p} \\ &= [h_{|[v_1;v_2]}]_{\mathbb{C}_p} + [f_{|[u_1;u_2]}]_{\mathbb{C}_p} \\ &= 0\end{aligned}$$




$f_{|[0;u_1]}$, $f_{|[u_1;u_2]}$ and $f_{|[u_2;1]}$ are respectively drawn in red, green and blue.

Figure 10: Prototypical examples of the cases of Corollary 3

□

inria-00564552, version 1 - 9 Feb 2011

*4.5. The proof of Theorem 1*

This gives us a Theorem that characterizes the arcs that go from a point $A$ to a point $B$ such that $B-A=1$ (*resp.* $-1$) and that do not go through $C$ and $D$ different from $A$ and $B$ so that $D-C=1$.

**Theorem 4** *The following statements are equivalent:*

1. *$f$ is a simple arc of the plane from a point $A$ to a point $B$ such that $B-A=1$ (resp. $-1$) and there is not $t_1$, $t_2$ such that $t_1<t_2$, $[t_1;t_2]\subsetneq[0;1]$ and $f(t_2)-f(t_1)=1$ (resp. $f(t_2)-f(t_1)=-1$).*
2. *$\exp\circ f$ is a Jordan curve in $\mathbb{C}_0$ whose winding number is 1 (resp. $-1$).*

*Proof.* We only treat the case where $B-A=1$, the other case being symmetric.

Let $g$ be the curve $\exp\circ f$. Clearly $g$ must have winding number 1, and, from Corollary 3, there are $v_1$ and $v_2$ such that $v_1<v_2$, $[v_1;v_2]\subsetneq[0;1]$ and $g(v_1)=g(v_2)$ if and only if it there are $t_1$ and $t_2$ such that $t_1<t_2$, $[t_1;t_2]\subsetneq[0;1]$ and $g_{|[t_1;t_2]}$ has 0 or 1 as winding number. In case $g_{|[t_1;t_2]}$ has winding number 0 then $f(t_1)=f(t_2)$ and $f$ is not simple. In case $g_{|[t_1;t_2]}$ has winding number 1, $f(t_2)-f(t_1)=1$. Thus if $g$ is not simple, $f$ is either not simple or there are $t_1$ and $t_2$ such that $t_1<t_2$, $[t_1;t_2]\subsetneq[0;1]$ and $f(t_2)-f(t_1)=1$.

Conversely, if $f$ is not simple or if there are $t_1$ and $t_2$ such that $t_1<t_2$, $[t_1;t_2]\subsetneq[0;1]$ and $f(t_2)-f(t_1)=1$ then clearly $g$ cannot be simple. □

In what follows $\rho_1$ is the inclusion mapping from $\mathbb{C}_{0,1}$ into $\mathbb{C}_0$ while $\rho_2$ is the inclusion mapping from $\mathcal{C}_{0,1}$ into $\mathcal{C}_1$.

**Theorem 5** *If $f$ is a Jordan curve in $\mathbb{C}_\mathbb{Z}$ if there are $k$ elements of $\mathbb{Z}$ in the interior of $f$ then the winding number of $\rho_2\circ e\circ f$ is either $k$ or $-k$.*

*Proof.* If $w$ is a word on $\{\alpha_k,\alpha_k^{-1}|k\in\mathbb{Z}\}^*$ representing $[f]_{\mathbb{C}_\mathbb{Z}}$, then

$$\widehat{\rho}_2(\widehat{\exp'}([f]_{\mathbb{C}_\mathbb{Z}}))=\sum_{k\in\mathbb{Z}}|w|_{\alpha_k}-|w|_{\alpha_k^{-1}}$$

where $|w|_{\alpha_k^\epsilon}$ is the number of occurrences in $w$ of $\alpha_k^\epsilon$. We also have that $\widehat{\iota}_k([f]_{\mathbb{C}_\mathbb{Z}})=\widehat{\iota}_k(w)=|w|_{\alpha_k}-|w|_{\alpha_k^{-1}}$. But

in $\mathbb{C}_k$. Then, from Lemma 6, we have that either for every $k$, $\widehat{\iota}_k([f]_{P_\mathbb{Z}})$ belongs to $\{0;1\}$ or for every $k$, $\widehat{\iota}_k([f]_{\mathbb{C}_\mathbb{Z}})$ belongs to $\{-1;0\}$. In the first case, we know that the winding numbers of $f$ around each element of $\mathbb{Z}$ in the interior of $f$ is 1 which implies that $\widehat{\rho}_2(\widehat{\exp'}([f]_{\mathbb{C}_\mathbb{Z}}))=k$. In the second case we similarly get that $\widehat{\rho}_2(\widehat{\exp'}([f]_{\mathbb{C}_\mathbb{Z}}))=-k$. □

The previous Theorem immediately implies the following corollary.

**Corollary 4.** *If $f$ is a Jordan curve in $\mathbb{C}_\mathbb{Z}$ so that 0 and 1 are in the interior of $f$ then $\rho_2\circ e\circ f$ is $k$ has a winding number that is either greater or equal to 2 or lower or equal to $-2$.*

inria-00564552, version 1 - 9 Feb 2011

Theorem 4 and Corollary 4 are enough to prove Theorem 1.

**Theorem 1** *If $f$ is a Jordan curve of $\mathbb{C}$ such that $f(0) = A$ and $f(t) = D$ for some $t$ in $(0;1)$ and there are two points $A'$ and $D'$ in the interior of $f$ such that $D - A = D' - A'$, then there is $t_1$ and $t_2$ such that $t_1 < t_2$ and one of the following property holds:*

1. $[t_1; t_2] \subsetneq [0; t]$ *and* $f(t_2) - f(t_1) = D - A$
2. $[t_1; t_2] \subsetneq [t; 1]$ *and* $f(t_2) - f(t_1) = A - D$

*Proof.* Without loss of generality, we may assume that $A' = 0$ and $D' = 1$. For any homotopic deformation of $f$ in $\mathbb{C}_{0,1}$ into a Jordan curve $f'$ that lies in $\mathbb{C}_{\mathbb{Z}}$, we know from Corollary 4 that $[\rho_2 \circ \exp' \circ f']_{\mathbb{C}_1}$ is in $\{k \mid k \leq -2 \vee k \geq 2\}$. While, if we suppose that $f$ does not fulfill the conclusion of the Theorem then, from Theorem 4, we get that the curve of $g = \exp \circ f$ is so that $g_1 = g_{|[0;t]}$ and $g_2 = g_{|[t;0]}$ are Jordan curves of $\mathbb{C}_0$ and that their respective winding numbers in $\mathbb{C}_0$ are 1 and $-1$.

In case $g$ lies in $\mathbb{C}_{0,1}$, *i.e.* there is no $t'$ such that $g(t') = 1$, Lemma 6 implies that the respective winding numbers of $g_1$ and $g_2$ in $\mathbb{C}_1$ are respectively in $\{0; 1\}$ and $\{-1; 0\}$ so that, in $\mathbb{C}_1$, $g$ has a winding number in $\{-1; 0; 1\}$. But we have that $[\rho_2 \circ \exp' \circ f]_{\mathbb{C}_1} = [\rho_2 \circ g]_{\mathbb{C}_1}$ and since, in such a case $f$ lies in $\mathbb{C}_{\mathbb{Z}}$ we get a contradiction with the fact that $[\rho_2 \circ \exp' \circ f]_{\mathbb{C}_1}$ should be in $\{k \mid k \leq -2 \vee k \geq 2\}$.

In case $g^{-1}(1) \neq \emptyset$, we are going to deform $g$ with an homotopy into a curve $g'$ so that:

1. $g'^{-1}(1) = \emptyset$
2. with the homotopy-lifting property, this deformation corresponds to a deformation of $f$ in $\mathcal{C}_{0,1}$

which will allow us to obtain a contradiction similarly to the previous case. There are four possibilities: only $g_1$ goes through 1, only $g_2$ goes through 1, $g_1(1) = g_2(0) = 1$, and $g_1$ and $g_2$ go both through 1 with $g_1(1) \neq 1$ and $g_2(0) \neq 1$. We are only going to treat this last case (the case $g_1(1) = g_2(0) = 1$ is easy and the other cases can be done in a similar way). This means that there is $t_1$ such that $f_{[|[0;t]]}(t_1) = k_1$ in $\mathbb{Z}$ and there is $t_2$ such that $f_{|[t;1]}(t_2) = k_2$ in $\mathbb{Z}$. Since $f$ is a Jordan curve, we must have that $k_1 \neq k_2$. Let $V$ be an open neighborhood of 1 that is evenly covered by exp, we let $W_1$ be the intersection of $\exp^{-1}(V)$ with an open neighborhood of $k_1$. Obviously $V_1 = \exp(W_1)$ is also evenly covered by exp, we can then perform an homotopy on the part of $g_1$ that lies in $V_1$ in order to obtain a Jordan curve $g'_1$ that does not go through 1. Similarly, we can obtain $W_2$ that is the intersection of $\exp^{-1}(V)$ with an open neighborhood of $k_2$, note that $W_1$ and $W_2$ are disjoint and cannot contain either 0 or 1. We can then obtain a Jordan curve $g'_2$ that does not go through 1 by deforming the part of $g_2$ that lies in $V_2 = \exp(W_2)$ with a homotopy. Let $g'$ be $g'_1 \cdot g'_2$, we call $f'$ the unique curve of $\mathbb{C}_{\mathbb{Z}}$ that can be obtained with unique path-lifting property of exp applied to $g'$ and so that $f'(0) = f(0)$. Using the homotopy-lifting property we have that $f'$ is homotop to $f$ in $\mathbb{C}_{0,1}$, furthermore

inria-00564552, version 1 - 9 Feb 2011

$f'$ lies in $\mathbb{C}_{\mathbb{Z}}$. It remains to show that $f'$ is a Jordan curve. Since $g'_1$ and $g'_2$ are both Jordan curves, we must have that $f'_{|[0;t]}$ and $f'_{|[t;1]}$ are both Jordan arcs. Thus if $f'$ is not a Jordan curve it must be from the fact that $f'_{|[0;t]}$ intersects $f'_{|[t;1]}$, but, from the homotopy lifting property $f'_{|[0;t]}$ is obtained from $f_{|[0;t]}$ by a homotopy that deforms it only in $W_1$ while $f'_{|[t;1]}$ is obtained from $f_{|[t;1]}$ by a homotopy that deforms it only in $W_2$ and as $W_1$ and $W_2$ are disjoint, it cannot be the case that the two arcs intersect each other. Thus, $f'$ is a Jordan curve that is homotop to $f$ in $\mathbb{C}_{0,1}$ while $[\rho_2 \circ \exp' \circ f']_{\mathbb{C}_1} \in \{-1; 0; 1\}$ which leads to a contradiction.

□

## 5. Conclusion

We proved that $MIX$ is a 2-MCFL, so that under the assumption that the class MCFLs capture the notion of mildly context sensitive languages, this proves that $MIX$ is a mildly context sensitive language. Since mildly context sensitive languages, as Joshi described them, should exclude *pathological* languages such as $MIX$, this implies that the class of formal languages that captures mildly context sensitivity should be explored again. A natural candidate would be *well-nested Multiple Context Free Languages* (MCFL$_{wn}$) as defined in [Kan09b]. The interest of MCFL$_{wn}$ lies in the fact that they satisfy a strong form of pumping Lemma as proved in [Kan09b] while there is still no strong form of pumping Lemma for MCFL [KS07]. Thus the class of MCFL is quite challenging in terms of open problems. Furthermore, the grammar we propose to capture $O_2$ (and the one we would deduce for $MIX$) strongly relies on non-well-nested rules, and it is easy to see that removing non-well-nested rules weaken the generative capacity of the grammar (for example when removing this rule the word $aabb\bar{a}b\bar{a}bbaaa\bar{b}\bar{b}b\bar{a}\bar{a}\bar{a}b$ is no more in the language generated by the grammar). Thus we conjecture that neither $MIX$ nor $O_2$ is a MCFL$_{wn}$. If our conjecture is right, and if we stick to the idea that $MIX$ should not be a mildly context sensitive language, then MCFL$_{wn}$ seem to be a possible replacement for the capturing mildly context sensitivity. But on the other hand, the result that $MIX$ is an MCFL shows that MCFL can be used to model certain scrambling phenomena. It is even known that if, for every $k$, $O_k$ (as defined in [FR68]) is an MCFL then this class should contain all the permutation closures of recognizable languages [Lat79]. According to [GS71], this would give a lot of control on scrambling phenomena (*e.g.* specify islands of scrambling). But the problem of showing whether $O_3$ is an MCFL appear to be quite challenging. Indeed, the proof we give strongly relies on the Jordan curve Theorem, *e.g.* for the definition of the invariant that the curve representation of words satisfy in the case 5, for Lemma 16 and in Theorem 5. Thus before trying to obtain any result on this problem, it seems reasonable to try and find an alternate proof that $O_2$ is an MCFL that would avoid the detour through geometry.

On the side of the word problem for $\mathbb{Z}^2$, it is still open whether it can be solved by indexed grammars. We conjecture it is not the case. Indeed, MCFL$_{wn}$

inria-00564552, version 1 - 9 Feb 2011

can be seen as non-duplicating macro languages as introduced by [Fis68] which are properly contained in indexed languages which are actually macro languages using the OI mode of derivation. It seems to us that semilinear rational cones (or semilinear Full Trio) that are included in indexed languages should be included in the class of $\text{MCFL}_{wn}$. And as the rational cone generated by $O_2$ is semilinear, if our first conjecture and our intuition about semilinear rational cones included in the class of indexed languages are correct, then it would follow that neither $O_2$ nor $MIX$ are indexed languages.

**Acknowledgement:** I would like to thank Philippe de Groote, Makoto Kanazawa and Géraud Sénizergues for fruitful discussions, their support and their patience when trying to understand my confusing explanations about curves. I wish also to thank Gilles Zémor for pointing to me the possible use of algebraic topology to cope with the problem on Jordan curves.

## References

[AdMO05] K. Aehlig, J.G. de Miranda, and C.-H.L. Ong. Safety is not a restriction at level 2 for string languages. In Vladimiro Sassone, editor, *Foundations of Software Science and Computational Structures*, volume 3441 of *Lecture Notes in Computer Science*, pages 490–504. Springer Berlin / Heidelberg, 2005.

[Ani71] A. V. Anisimov. Group languages. *Cybernetics and Systems Analysis*, 7(4):594–601, 1971.

[AS60] Lars V. Ahlfors and Leo Sario. *Riemann Surfaces*, chapter 1. Princeton University Press, 1960.

[Bac81] Emmon Bach. Discontinuous constituents in generalized categorial grammars. In Victoria Burke and James Pustejovsky, editors, *Proceedings of the llth Annual Meeting of the Northeastern Linguistics Society*, pages 1–12, 1981.

[Bac88] Emmon Bach. Categorial grammars as theories of language. In Richard T. Oehrle, Emmon Bach, and Deirdre Wheeler, editors, *Categorial Grammars and Natural Language Structures*, pages 17–34. D. Reidel, 1988.

[Boo58] William W. Boone. The word problem. In *Proceedings of the National Academy of Sciences of the Unites States of America*, pages 1061–1065, 1958.

[Dam82] Werner Damm. The IO- and OI-hierarchies. *Theoretical Computer Science*, 20:95–207, 1982.

[DG86] Werner Damm and Andreas Goerdt. An automata-theoretical characterization of the OI-hierarchy. *Information and Control*, 71(1-2):1–32, 1986.

inria-00564552, version 1 - 9 Feb 2011

[dGP05] Philippe de Groote and Sylvain Pogodalla. On the expressive power of abstract categorial grammars: Representing context-free formalisms. *Journal of Logic, Language and Information*, 13(4):421–438, 2005.

[Fis68] Michael J. Fischer. *Grammars with macro-like productions*. PhD thesis, Harvard University, 1968.

[FR68] Michael J. Fischer and Arnold L. Rosenberg. Real-time solutions of the origin-crossing problem. *Mathematical Systems Theory*, 2(3):257–263, 1968.

[GGH69] S. Ginsburg, S. Greibach, and J. Hopcroft. *Studies in abstract families of languages*. Number 87 in Memoirs. American Mathematical Society, 1969.

[GO90] Bernard R. Gelbaum and John M.H. Olmsted. *Theorems and Counterexamples in Mathematics*, chapter 2, pages 42–185. Problem Books in Mathmatics. Springer, 1990.

[GS71] Seymour Ginsburg and Edwin H. Spanier. Afl with the semilinear property. *J. Comput. Syst. Sci.*, 5(4):365–396, 1971.

[HMOS08] Matthew Hague, Andrzej S. Murawski, C.-H. Luke Ong, and Olivier Serre. Collapsible pushdown automata and recursion schemes. In *LICS*, pages 452–461. IEEE Computer Society, 2008.

[Jos85] Aravind K. Joshi. Tree-adjoining grammars: How much context sensitivity is required to provide reasonable structural descriptions? In David Dowty, Lauri Karttunen, and Arnold M. Zwicky, editors, *Natural Language Parsing*, pages 206–250. Cambridge University Press, 1985.

[JSW91] Aravind K. Joshi, Vijay K. Shanker, and David J. Weir. The converence of mildly context-sensitive grammar formalisms. In Peter Sells, Stuart M. Shieber, and Thomas Wasow, editors, *Foundational Issues in Natural Language Processing*, pages 31–81. The MIT Press, 1991.

[Kan09a] Makoto Kanazawa. The convergence of well-nested mildly context-sensitive grammar formalisms, July 2009.

[Kan09b] Mokoto Kanazawa. The pumping lemma for well-nested multiple context-free languages. In V. Diekert and D. Nowotka, editors, *proceedings of DLT 2009*, volume 5583 of *LNCS*, pages 312–325. Springer, 2009.

[KNU02] Teodor Knapik, Damian Niwinski, and PawełUrzyczyn. Higher-order pushdown trees are easy. In Springer, editor, *Proc. FoSSaCS'02*, volume 2303, pages 205–222, 2002.

inria-00564552, version 1 - 9 Feb 2011

[KS07] Makoto Kanazawa and Sylvain Salvati. Generating control languages with abstract categorial grammars. In Laura Kallmeyer, Paola Monachesi, Gerald Penn, and Giorgio Satta, editors, *Proceedings of the 12th conference on Formal Grammar (FG 2007), Dublin, Ireland*. CSLI Publications, August 2007. To appear, ISSN 1935-1569.

[KS10] Makoto Kanazawa and Sylvain Salvati. The copying power of well-nested multiple context-free grammars. In Adrian-Horia Dediu, Henning Fernau, and Carlos Martín-Vide, editors, *Language and Automata Theory and Applications*, volume 6031, pages 344–355. Lecture Notes in Computer Science, 2010.

[Laf09] Yves Lafont. Réécriture et problème du mot. *Gazette des mathématiciens*, 120:27–38, 2009.

[Lat79] Michel Latteux. Cônes rationnels commutatifs. *J. Comput. Syst. Sci.*, 18(3):307–333, 1979.

[Mic09] Jens Michaelis. An additional observation on strict derivational minimalism. In James Rogers, editor, *Proceedings of FG-MoL 2005: The 10th conference on Formal Grammar and the 9th Meeting on Mathematics of Language*, pages 101–111. CSLI publications, 2009.

[MS83] David E. Muller and Paul E. Schupp. Groups, the theory of ends, and context-free languages. *Journal of Computer and System Sciences*, 26(3):295–310, 1983.

[Niv68] Maurice Nivat. *Transductions des langages de Chomsky*. Thèse d'état, Annales de l'institut Fourier, 1968.

[Nov55] Pyotr Sergeyevich Novikov. On the algorithmic unsolvability of the word problem in group theory. *Trudy Mat. Inst. Steklov*, 44:1–143, 1955.

[Pul83] Geoffrey K. Pullum. Context-freeness and the computer processing of human languages. In *Proceedings of the 21st annual meeting on Association for Computational Linguistics*, pages 1–6. Association for Computational Linguistics, 1983.

[Sal07] Sylvain Salvati. Encoding second order string acg with deterministic tree walking transducers. In S. Wintner, editor, *Proceedings FG 2006: the 11th conference on Formal Grammars*, FG Online Proceedings, pages 143–156. CSLI Publications, 2007.

[SMFK91] Hiroyuki Seki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. On multiple context free grammars. *Theoretical Computer Science*, 88(2):191–229, 1991.

inria-00564552, version 1 - 9 Feb 2011

[Spa81] Edwin H. Spanier. *Algebraic Topology*. Springer-Verlag, corrected reprint edition, 1981.

[Wei88] David J. Weir. *Characterizing mildly context-sensitive grammar formalisms*. PhD thesis, University of Pennsylvania, Philadephia, PA, 1988. Supervisor-Aravind K. Joshi.

[Wei92] David J. Weir. Linear context-free rewriting systems and deterministic tree-walking transducers. In *ACL*, pages 136–143, 1992.