# Formal Languages & Word-Rewriting

Géraud Sénizergues

LaBRI
(the laboratory at Bordeaux for research in computing)
Université de Bordeaux I
**

## 1 INTRODUCTION

In this paper, we present results that link the theory of word-rewriting (also known as the theory of *semi-Thue* systems) to the theory of formal languages.

Thue systems were first defined by A. Thue in 1914 ([Thu14]). Since then, these systems have been the subject of much research with respect to decidability problems dealing with the relations $\xrightarrow[S]{*}, \xleftrightarrow[S]{*}$, produced by such systems. For examples, see the contributions of Y. Lafont and A. Prouté [LP94], G. Lallement [Lal94] as well as Y. Matiyasevich [Mat94] in this volume.

Formal language theory has, as its objects, the sets of words written over an alphabet X. This theory is concerned primarily with classifying these sets either according to the method of generating them (such methods as grammars, iterated morphisms, etc.) or according to the means of recognizing them (such means as automata, morphisms in algebraic structure, logical formulae, etc.).

A semi-Thue system obviously offers a way of generating words: we could define a language $L \subset X^*$ as the set of words derived from an axiom $u \in X^*$ modulo a semi-Thue system S. In consequence, studying the links between this way of defining a language and the conventional ways we just mentioned seems like a natural pre-occupation in language theory.

In fact, Maurice Nivat had already brought up this area of study in the '70s and had even pushed forward the first systematic work on this topic ([Ben69, Niv70, Niv71, CN71, Coc71, NB72, But73b, But73a, Coc75, Coc76, Ber77] [3]). Since that time, all sorts of researchers have contributed to this area, studying the links between semi-Thue systems and formal languages. ([AB92, BS86, Ben87, Boa80, BN84, BO85, Boo87, BJW82, BO93, Car91, Cau88, Cau89, Cau93,

** mailing address:LaBRI & UER Math-info, Université Bordeaux I
351, cours de la Libération
33405 Talence Cedex
France
e-mail:ges@labri.u-bordeaux.fr
fax: (+33) 56-84-66-69

[3] [Büc64, Gre67],and [McN67] take off from this idea; [McN67] is cited in [Niv70] as a guideline

[4] Even this list of works is far from exhaustive. [Jan88, BO93], for example, is also useful.

Cho79, Cho82, Jan88, Kie79, MNOZ93, McC93, MNO88, NOR86, Nar86, Nar90, O'D81, O'D83, Ott84, Ott87, OZ91, Ott92a, Ott92b, Sak79, Sén81, Sén85, Sén86, Sén87, Sén89, Sén90a, Sén90b, Zha92] [4])

In line with these antecedents, the theorems presented in this paper will take the following schematic form:

**Characterization Theorem** : Let $L$ be a language. The following properties are equivalent:

(1) $L$ is in the class $\mathcal{L}$.

(2) There exists a semi-Thue system $S$ of the class $\mathcal{S}$ and a word $f$ such that

$$L = [f]_{\underset{S}{\overset{*}{\longleftrightarrow}}}$$

When we have only (1) $\Longrightarrow$ (2), we speak of a "theorem about *representation* of the languages of $\mathcal{L}$."

When we have only (2) $\Longrightarrow$ (1), we speak of a "theorem about the *structure* of congruence classes of the systems of $\mathcal{S}$".

These preliminaries provide a new means of investigating the properties of semi-Thue systems or of formal languages, as we can see from the initial work of [Büc64, Gre67, McN67]. After each characterization theorem, we will sytematically cite *applications* to word-rewriting or to formal languages.

## 2 PRELIMINARIES

### 2.1 Abstract reductions

Let $E$ be some set and $\longrightarrow$ some binary relation over $E$ .We shall call $\longrightarrow$ the *direct reduction*.We shall use the notations $\xrightarrow{i}$ for every integer $(i \geq 0)$, and $\xrightarrow{*}, \xrightarrow{+}$ in the usual way (see [Hue80]). The inverse relation , $\longrightarrow^{-1}$ will be denoted by $\longleftarrow$ . By $\longleftrightarrow$, we denote the relation $\longrightarrow U \longleftarrow$. The three relations $\xrightarrow{*}, \xleftarrow{*}$ and $\overset{*}{\longleftrightarrow}$ are respectively the reduction, derivation and the equivalence generated by $\longrightarrow$.

We shall use the notions of *confluent* reduction, *locally confluent* reduction, *Church-Rosser* reduction and *noetherian* reduction in their usual meaning ([Hue80] or [DJ91, §4, p.266-269]).

An element $e \in E$ is said *irreducible* modulo $(\longrightarrow)$ iff there exists no $e' \in E$ such that $e \longrightarrow e'$. By Irr$(\longrightarrow)$ we denote the set of all the elements of $E$ which are irreducible modulo $(\longrightarrow)$.

An element $e \in E$ is said $\longrightarrow$-reducible iff, there exists some $e' \in E, e \longrightarrow e'$.

Given some subset $A$ of $E$, we use the following notations:

$$< A >_{\xleftarrow{*}} = \{e \in E \mid \exists a \in A, a \xleftarrow{*} e\}$$

$$[A] \overset{*}{\longleftrightarrow} = \{e \in E \mid \exists a \in A, a \overset{*}{\longleftrightarrow} e\}$$

The relation $\longrightarrow$ is said *partly confluent* over $A$ if and only if :

$$< A > \overset{*}{\longleftarrow} = [A] \overset{*}{\longleftrightarrow}$$

(which is equivalent to: $\forall a \in A, \forall e \in E$, if $a \overset{*}{\longleftrightarrow} e$ then $\exists a' \in A$ such that $a' \overset{*}{\longleftarrow} e$)

## 2.2 Reductions over a free monoïd

Let us consider now the particular case where $E = X^*$, the free monoïd generated by some alphabet $X$. $\longrightarrow$ is said *l-decreasing* (resp. *strictly l-decreasing*) iff it reduces (resp. strictly reduces) the length, i.e.:

$$\forall f \in X^*, \forall g \in X^*, f \longrightarrow g \Longrightarrow \mid f \mid \geq \mid g \mid (resp. \mid f \mid > \mid g \mid).$$

We shall abreviate "strictly l-decreasing" by "strict".

A *valuation* over $X^*$ is an homomorphism $\nu : (X^*, .) \longrightarrow (\mathbb{N}, +)$ such that, for every $x \in X, \nu(x) \neq 0$.

$\longrightarrow$ is said *v-decreasing* (resp. *strictly v-decreasing*) iff, there exists some valuation $\nu$ over $X^*$, such that $\longrightarrow$ reduces (resp. strictly reduces) the valuation $\nu$, i.e.:

$$\forall f \in X^*, \forall g \in X^*, f \longrightarrow g \Longrightarrow \nu(f) \geq \nu(g) (resp. \nu(f) > \nu(g))$$

We shall abreviate "strictly v-decreasing" by "v-strict".

## 2.3 Controlled rewriting systems

The notion of *controlled rewriting system* over an alphabet $X$ can be seen as generalising the notion of semi-Thue system over $X$: to every rule $u \longrightarrow v$ is associated some set of words $K(u,v)$ which is the set of left-contexts with which the given rule can be used: a word $pus$ can be rewritten $pvs$ only when $p \in K(u,v)$ [5].

In the particular case where for every rule $u \longrightarrow v, K(u,v)$ is the whole free monoïd $X^*$ , one recovers the classical notion of semi-Thue system.

Let us give now formal definitions.

[5] hence we should call *left*-controlled rewriting systems such systems;for sake of brevity we have dropped the prefix "left"

A controlled rewriting system over the alphabet $X$ is a subset $S$ of $X^* \times X^* \times X^*$. Every element $(l,u,v)$ of $S$ is a *rule* of $S$. $u$ is the *lefthand side* and $v$ the *righthand side* of the rule $(l,u,v)$. By LH($S$) (resp. RH($S$)) we denote the set of lefthand (resp. righthand) sides of rules of $S$. The direct reduction generated by $S$ (which is denoted by $\underset{S}{\longrightarrow}$) is defined by :

for every $f,g \in X^*, f \underset{S}{\longrightarrow} g$ iff there exists $(l,u,v) \in S$ and $s \in X^*$ such that $f = lus$ and $g = lvs$.

The relations $\overset{*}{\underset{S}{\longrightarrow}}$ (*reduction generated by* $S$), $\overset{*}{\underset{S}{\longleftarrow}}$ (*derivation generated by* $S$) and $\overset{*}{\underset{S}{\longleftrightarrow}}$ (*equivalence generated by* $S$) are then fully defined from $\underset{S}{\longrightarrow}$ by the general definitions given in §2.1. One can check that $\overset{*}{\underset{S}{\longleftrightarrow}}$ is the smallest right-congruence over $(X^*,.)$ which contains $\{(lu,lv) \mid (l,u,v) \in S\}$. $\overset{*}{\underset{S}{\longleftrightarrow}}$ is then called the *right-congruence generated by* $S$.

## 2.4 Classes of controlled rewriting systems

Let $\mathcal{C}_1, \mathcal{C}_2$ be two classes of languages. Let us denote by $\mathcal{C}_i(X)$ the set of languages over the alphabet $X$ which belong to the class $\mathcal{C}_i$ (for $i \in \{1,2\}$).

We call $\mathcal{C}_1 - \mathcal{C}_2$*-decomposition* over $X$ every finite set

$$D = \{L_i \times V_i \times \{u_i\}\}_{i\in[1,n]}$$

such that,

$$\forall i \in [1,n], L_i \in \mathcal{C}_1(X), V_i \in \mathcal{C}_2(X), u_i \in X^*.$$

We call *component* of $D$ every element $L_i \times V_i \times \{u_i\}$ of $D$.

With every decomposition $D$ is associated a controlled rewriting system $\dot{D}$ as follows:

$$\dot{D} = \bigcup_{i=1}^{n} L_i \times V_i \times \{u_i\}$$

We say that some controlled rewriting system $S$ belongs to the class $\mathcal{C}_1 - \mathcal{C}_2$ iff, there exists some $\mathcal{C}_1 - \mathcal{C}_2$-decomposition $D$ such that $\dot{D} = S$. We shall also say that $S$ is a $\mathcal{C}_1 - \mathcal{C}_2$ controlled rewriting system.

Let us mention below some classes of controlled rewriting systems which can be found in the litterature. Let us denote by Rec, Cf , Det, Rat, Fin respectively the classes of recursive, context-free, deterministic context-free, rational, finite languages and let us abreviate "controlled rewriting system" by "c-system".

The notion of Rec-Fin c-system was introduced in [But73a] in order to devise an equivalence algorithm for the so-called "simple" grammars , which, as in the case of "very-simple" grammars ([But73b]), would lean on the comparison of finitely generated congruences.

It was noticed in [Cho79] that the c-systems considered in [But73a] belong to the class Det-Fin. The notion of Rat-Fin c-system is studied in [Cho79, Cho82, Kie79], in connection with the notions of context-free or deterministic context-free language. The rational semi-Thue systems , studied in [O'D81, O'D83, Nar86] can be considered as particular cases of Rat-Rat c-systems. As well, the context-free semi-Thue systems, studied in [BJW82] are particular cases of Rat-Alg c-systems. A result of [BN84] shows that every Fin-Rat c-system defines a reduction which is a rational transduction (we state this in full details in §3.1). In the following, we call *finite* c-system every c-system in the class Rat-Fin. As well, we call *finite* decomposition, every Rat-Fin-decomposition.

## 2.5 Combinatorial properties of controlled rewriting systems

Let $S$ be some c-system. $S$ is said confluent, locally confluent, Church-Rosser, noetherian, strictly l-decreasing, l-decreasing, strictly v-decreasing, v-decreasing if and only if, the relation $\underset{S}{\longrightarrow}$ fulfills that property. We use the notation $Irr(S)$ in place of $Irr(\underset{S}{\longrightarrow})$ and we write that some word $f$ is $S$-irreducible (resp. $S$-reducible) in place of $\underset{S}{\longrightarrow}$-irreducible (resp. $\underset{S}{\longrightarrow}$-reducible).

Let us now define some *combinatorial* properties which depend really on the c-system $S$ itself (and not on the reduction $\underset{S}{\longrightarrow}$ only).

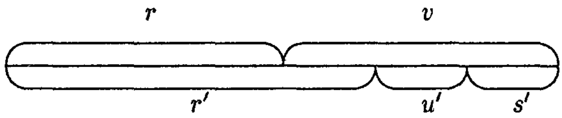
**Definition** : Let us consider the following conditions on the rules of some c-system $S$ :

C1 : for every $(r, u, v), (r', u', v') \in S$ and every $s' \in X^*$
$rv = r'u's'$ and $\mid r \mid \leq \mid r' \mid \Longrightarrow \mid s' \mid = 0$ and $\mid r \mid = \mid r' \mid$
C2 : for every $(r, u, v), (r', u', v') \in S$ and every $s' \in X^*$
$rv = r'u's'$ and $\mid r \mid > \mid r' \mid \Longrightarrow \mid s' \mid = 0$ or $\mid r'u' \mid \leq \mid r \mid$
C3 : for every $(r, u, v), (r', u', v') \in S$ and every $s \in X^*$
$rvs = r'u's'$ and $\mid r \mid < \mid r' \mid \Longrightarrow \mid rv \mid \leq \mid r' \mid$
$S$ is said *left-basic* iff it fulfills C1 and C2
$S$ is said *right-basic* iff it fulfills C1 and C3
$S$ is said *basic* iff it fulfills C1, C2 and C3 (which is equivalent to say that $S$ is both left- and right- basic).

Each condition $C_i (i \in [1, 3])$ can be seen as the prohibition of some superposition configuration for two-rules $(r, u, v), (r', u', v')$ of $S$.

**Condition C1** : C1 expresses the prohibition of the following configuration: $rv = r'u's'$ where $\mid r \mid \leq \mid r' \mid$ and $\mid u' \mid < \mid v \mid$

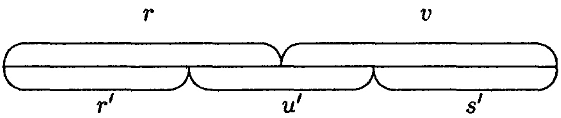
schema 1:




In other words, a righthand side of rule may not strictly embed any lefthand side of rule.

**Condition C2** : C2 expresses the prohibition of the following configuration: $rv = r'u's'$ where $0 <| s' |$ and $| r' |<| r |<| r'u' |$

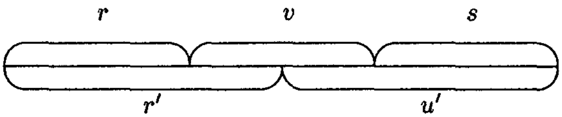
schema 2:




In other words, a righthand side of rule may not be "strictly overlapped on the left" by any lefthand side of rule.

**Condition C3** : C3 expresses the prohibition of the following configuration: $rvs = r'u'$ where $| r |<| r' |<| rv |$.

schema 3:




In other words, a righthand side of rule may not be "strictly overlapped on the right" by any lefthand side of rule.

These definitions were given in [Niv70, Coc71, But73b, Sak79] in the case of semi-Thue systems and were generalised to controlled systems in [But73a, Cho79]. This notion of *basic* system turns out to be a key notion for the study of the links between context-free langages and finitely generated congruences (or right-congruences generated by a finite c-system).

## 2.6 Finite & rational semi-Thue systems

A semi-Thue system over the alphabet $X$ is ( by definition) some subset $T$ of $X^* \times X^*$. When the set $T$ is finite , one says that $T$ is a *finite* semi-Thue system.

A semi-Thue system $T$ over $X$ is said *rational* ([O'D81, O'D83]) iff it satisfies some equality :

$$T = \bigcup_{i=1}^{n} V_i \times \{u_i\}$$

where $n$ is some integer, each $V_i$ is some rational language over $X$ and each $u_i$ is some word over $X$. With every semi-Thue system $T$ over $X$ one can naturally associate a c-system $\overline{T}$ over $X$ by:

$$\overline{T} = \{(l, u, v) \in X^* \times X^* \times X^* \mid (u, v) \in T\}$$

or, in other words $\overline{T} = X^* \times T$. It is clear that the direct reductions $\underset{\overline{T}}{\longrightarrow}$ and $\underset{T}{\longrightarrow}$ are equal (where $\underset{\overline{T}}{\longrightarrow}$ is the relation defined in §2.3 and $\underset{T}{\longrightarrow}$ is the reduction generated by $T$, in the usual meaning).

Hence, one of both systems$\overline{T}, T$ is confluent (resp. locally confluent, Church-Rosser, noetherian, strict, l-decreasing, v-strict, v-decreasing) iff the other fulfills the same property.

$T$ will be said *left-basic* (resp. *right-basic, basic*) iff $\overline{T}$ is left-basic (resp. *right-basic, basic*) in the sense of definition 2.5 . This definition of the notion of basic semi-Thue system is equivalent to the definitions given in [Niv70] and [Ber77].

This definition of the notion of left-basic semi-Thue is equivalent, (up to some details) to the definition given in [Sak79].

A semi-Thue system $S$ is said:

- *monadic* iff $\forall (u, v) \in S, 1 \leq \mid u \mid$ and $\mid v \mid \leq 1$
- *special* iff $\forall (u, v) \in S, 1 \leq \mid u \mid$ and $\mid v \mid = 0$

One can notice that every monadic or special semi-Thue system is basic.

## 2.7 General references

Some *general* surveys on semi-Thue systems (they do *not* focus exclusively on the links with formal languages ) have been written successively in [Ber77, Boo87, Jan88, BO93].

In [BB91, §4, p. 95-100] and [MO89] one will find information on the links between semi-Thue systems , the so-called *context-free* groups and the *group languages*.We have left this subject out of this paper.

The language-theoretic notions used in this work are very classical and can be found in every treatise on the subject, for example [Aut87, Har78, HU79, KS86, Sal73].

# 3 RATIONAL LANGUAGES & WORD-REWRITING

## 3.1 Büchi systems

Büchi defined and studied in [Büc64, Büc90] what we shall call the *left-linear reduction* generated by a semi-Thue system $S$(he was seeing this kind of reductions as variants of the deduction relations generated by Post systems). The direct left-linear reduction generated by $S$, is denoted by $\xrightarrow[ll\,S]{}$ and is defined by: for every $f, g \in X^*, f \xrightarrow[ll\,S]{} g$ iff there exist $(u, v) \in S$ and $s \in X^*$ such that $f = us$ and $g = vs$.

The left-linear reduction relation, denoted by $\xrightarrow[ll\,S]{*}$, is then the reflexive and transitive closure of relation $\xrightarrow[ll\,S]{}$.

The *left-linear derivation* relation generated by $S$ is the inverse of the left-linear reduction and will be denoted by $\xleftarrow[S\,ll]{*}$. The *right-congruence* generated by $S$ , denoted by $\xleftrightarrow[ll\,S]{*}$,is the reflexive, symmetric and transitive closure of relation $\xrightarrow[ll\,S]{}$

**Theorem 3.1** ([Büc64]) *Let $L \in X^*$.The following properties are equivalent*

(1) $L =< F >_{\xleftarrow[S\,ll]{*}}$
*for some finite set $F \subset X^*$ and some finite semi-Thue system $S \subset X^* \times X^*$*
(2) $L =< F >_{\xleftarrow[S\,ll]{*}}$
*for some finite set $F \subset X^*$ and some finite semi-Thue system $S \subset X^* \times X^*$ of the form $S = \{(f_i p_i, f_i) | 1 \leq i \leq n\}$*
(3) *$L$ is rational*

**Complement 3.2** ([Coc75]) *Let $L \in X^*$.The following properties are equivalent*

(3) *$L$ is rational*
(4) $L = [F]_{\xleftrightarrow[ll\,S]{*}}$
*for some finite set $F \subset X^*$ and some finite semi-Thue system $S \subset X^* \times X^*$ such that $\xrightarrow[ll\,S]{}$ is strictly decreasing and confluent.*

Theorem 3.1 is used in [Gre67] in order to show that any relation "computed" by some pushdown-automaton preserves rationality.
The following theorem brings more precisions about the structure of the relation $\xrightarrow[\Pi_S]{*}$ itself.

**Theorem 3.3 ([BN84, Cau88])** *Let $S$ be some recognizable subset of $X^* \times X^*$. Then $\xrightarrow[\Pi_S]{*}$ is a rational transduction.*

This theorem 3.3 is the key argument in the proof of the fact that the "center"of any context-free language is context-free too ([BN84]).

### 3.2 Special systems

We recall that a *special* semi-Thue system over $X$ is any $S \subset X^+ \times \{\epsilon\}$. A subset $P$ of $X^*$ is said *unavoidable in* $X^*$ iff the language $X^* - X^*.P.X^*$ is finite.Some ordering $\preceq$ over some set $E$ is a *well* ordering iff, for every sequence $(e_i)_{0 \leq i}$ of elements of $E$, there exists a strictly increasing sequence of integers

$$i_1 < i_2 \ldots i_j < i_{j+1} < \ldots$$

such that

$$(e_{i_j})_{0 \leq j} \text{ is an increasing sequence with respect to } \preceq .$$

By $\alpha(S)$ we denote the set of letters appearing in the rules of system $S$.

**Theorem 3.4 ([EHR83])** *Let $S$ be a finite, special semi-Thue system over some finite alphabet $X$, let us suppose that $X = \alpha(S)$ and let $L = < \epsilon >_{\overset{*}{\underset{S}{\longleftarrow}}}$. Then the following properties are equivalent*

(1) *$L$ is rational*
(2) *$LH(S)$ is unavoidable in $X^*$*
(3) *$\overset{*}{\underset{S}{\longleftarrow}}$ is a well ordering over $X^*$*

An immediate corollary of theorem 3.4 is that rationality of a language of the form $< \epsilon >_{\overset{*}{\underset{S}{\longleftarrow}}}$, where $S$ is finite, special, is decidable. (Let us recall that every such language is context-free but the general problem whether some context-free language is rational is undecidable). Theorem 3.4 is used in [Sén86] where it is shown that every rational language in the free *group* with finite basis $X$, has a syntactic congruence which either has finite index (which is always the case in the free *monoïd*) either is the equality relation.

Given a semi-Thue system $S$ and a set of words $W$, by $\Delta_S^*(W)$ we denote the set of "descendants" of $W$ modulo $S$ i.e.:

$$\Delta_S^*(W) = \{v \in X^* | \exists w \in W, w \xrightarrow[S]{*} v\}$$

**Theorem 3.5** ([BS86]) *Let $S$ be a cancellation system , that is to say a finite semi-Thue system $S$ such that $\forall (u,v) \in S, |u| = 2, |v| = 0$. Then*

**(1)** *for every rational set $R$, $\Delta_S^*(R)$ is rational*
**(2)** *a finite automaton recognizing $\Delta_S^*(R)$ can be built in time $O(n^3)$ from any finite automaton with size $n$ recognizing $R$.*

The first result in this spirit was obtained in [Ben69];a first generalisation was given in [BO85] and theorem 3.5 improves on its complexity.
Remarks:

1. Point (1) of theorem 3.5 can be formulated as follows: the relation $\xrightarrow[S]{*}$ preserves rationality.
2. But relation $\xrightarrow[S]{*}$ is not a rational transduction in general: the languages $< R >_{\overset{*}{\underset{S}{\longleftarrow}}}$ with $R$ rational and $S$ cancellation system, are in general context-free but not rational
3. Conclusion (1) may fail as soon as $R$ is merely assumed context-free (see part 6)
4. Conclusion (1) may fail as soon as $S$ is merely assumed strictly decreasing (even when $\xrightarrow[S]{}$ is assumed confluent, see part 6)

**Complement 3.6**
**(1)** *If $S$ is assumed to be a finite,basic,strict,semi-Thue system, conclusion (1) remains valid and complexity (2) becomes at most $O(n^4)$ ([Ben87])*
**(2)** *If $S$ is assumed to be a finite,left-basic,strict,semi-Thue system, conclusion (1) remains valid ([Sak79])*
**(3)** *If $S$ is assumed to be a finite,left-basic,strict, c-system, conclusion (1) remains valid ([Sén90b])*
**(4)** *If $S$ is assumed to be a general semi-Thue system (maybe infinite, non-rational), which is basic, then conclusion (1) remains valid ([Sén81] in the general case, [BJW82] in the monadic case)*[6]

Theorem 3.5 allows one to prove that the family of rational subsets of the free group is closed under intersection and complement ([Ben69]). An application of theorem 3.5 to communication protocols is given in [BO85]. This theorem 3.5 is widely used in the proofs of the results exposed in part 4.

[6] but the properties "basic" and "monadic" though the first one generalises the second one, are essentially similar.

# 4 DETERMINISTIC CONTEXT-FREE LANGUAGES & WORD-REWRITING

## 4.1 Characterisations

The *deterministic context-free* languages are defined as the languages which can be recognized by some *deterministic* pushdown automaton ([GG66]).

**Theorem 4.1** *Let $L \subset X^*$. The following properties are equivalent*

**(1)** *$L$ is deterministic context-free*

**(2)** $L = [R]_{\overset{*}{\underset{S}{\longleftrightarrow}}}$

*where $R$ is some rational language over $X^*$ and $S$ is some strict,finite,basic,confluent, c-system*

**(3)** $\#L\$ = [f]_{\overset{*}{\underset{S}{\longleftrightarrow}}} \bigcap (X + \# + \$)^*$

*where $f$ is some word in $X^*$, $S$ is some v-strict,finite,left-basic, confluent semi-Thue system and $\#, \$$ are letters which do not belong to the alphabet $X$.*

Proof(s):

(1) $\Longrightarrow$ (2): is shown in [Sén90a]
(2) $\Longrightarrow$ (3): is shown in [Sén90b]
(3) $\Longrightarrow$ (1): is shown in [Sak79]

A direct proof of (2) $\Longrightarrow$ (1) is given in [Cho79] and [Cho82].

**Complement 4.2**

- *All the structure theorems (in the sense defined in our introduction)* [7] *asserting conclusion (1), follow from theorem 4.1 and, more precisely, from part* (2) $\Longrightarrow$ (1)
- *Every deterministic context-free language fulfills formula (3) for some semi-Thue system $S$ which is* strict, *finite and confluent ([MNO88]) (but it remains unknown whether $S$ can be choosen simultaneously* strict, *finite,* left-basic *and confluent)*

## 4.2 The equivalence problem for deterministic pushdown automata

Let us recall the *equivalence problem* for deterministic pushdown automata is the following decision problem: given two deterministic pushdown automata $A_1$,$A_2$, is it true that the associated languages $L(A_1)$,$L(A_2)$ are *equal*?

[7] excepted the tight result of [RS86]

Though a great deal of work has been devoted to this problem since the time it was raised ([GG66]), this problem remains open.We give below some applications of the results of the above section to this field of research.

**Application 4.3**
**(1)([Sén89])** *The following problem is decidable*
*instance:* $L_1 = [R]_{\overset{*}{\underset{S}{\longleftrightarrow}}}$ *and* $L_2 = L(A)$

*where S is some strict,rational,basic,confluent semi-Thue system and A is some general deterministic pushdown automaton.*
*question :* $L_1 = L_2$*?*

**(2)([Sén89]** *Let C be a family of languages which is an* effective cylinder[8]. *If the equivalence problem is decidable for pairs of languages* $(L_1, L_2)$ *both in this family, then the equivalence problem is decidable for pairs of languages* $(L_1, L_2)$ *where* $L_1$ *belongs to C and* $L_2$ *is some general deterministic context-free language.*
**(3)([Sén90b]** *The following problems are Turing-equivalent* [9]
**(P1)** *The equivalence problem for D.P.D.A.*
**(P2)** *The problem "*$[f]_{\overset{*}{\underset{S_1}{\longleftrightarrow}}} = [f]_{\overset{*}{\underset{S_2}{\longleftrightarrow}}}$ *?"*

*for* $S_1, S_2$ *strict,finite,left-basic,confluent semi-Thue systems*
**(P3)** *The problem "*$< f >_{\overset{*}{\underset{S}{\longleftarrow}}} = [f]_{\overset{*}{\underset{S}{\longleftrightarrow}}}$ *?"*

*for S strict,finite,left-basic semi-Thue system.*

Result (1) is obtained by means of a combination of theorem 4.1 with Valiant's theorem ([Val74],[Bee76]) [10].
Result (2) is obtained directly from theorem 4.1 by means of a notion of syntactic right-congruence "with left contexts" associated to language $L_1$. This result solves (in some way)the problem raised in [OIH81]: "is it possible to find some *general* scheme extending any equivalence algorithm for some subclass $\mathcal{A}$ of D.P.D.A. to an equivalence algorithm for pairs of D.P.D.A. *one of which* is in the subclass?".
Result (3) is obtained from theorem 4.1,point (2), by a sequence of transformations of rewriting systems ([Sén90b, pages 318-340]).

Other results have been obtained by means of rewriting systems (and are neither corollaries nor extensions of theorem 4.1):

[8] we call here *effective* cylinder any family of languages which is *effectively* closed by intersection with rational sets and inverse homomorphism, see [Sén89] for a precise definition

[9] they are even "many-one" equivalent but the proof quoted above does not establish this explicitly

[10] this theorem states that the equivalence problem is *decidable* for the subclass of *finite-turn* D.P.D.A

- The equivalence problem is decidable for *parenthetic* grammars ([McN67]), *nest-sets*([Tak75]) and more generally for *NTS* grammars ([Boa80, Sén85]) and even for *pré*-NTS grammars ([AB92]). This class of grammars and of generalisations is further studied in [Zha92, McC93].
- The best known algorithm testing the equivalence for *simple* grammars is given in [Cau89], it uses some finite, v-strict and confluent semi-Thue system which generates the congruence "to generate the same language" over the words written in the alphabet of the grammar; a generalisation to *stateless* automata is given in [Cau93].

### 4.3 Decision problems for rewriting systems

Theorem 4.1 and some of its applications (particularly 4.3) raise naturally some decision problems about rewriting systems.

**Theorem 4.4** *The confluence property is decidable for strict,finite,left-basic c-systems.*

This theorem was first shown by [O'D83] in the case of strict,rational, monadic semi-Thue systems and then generalised in [Sén90b] to the above case. Let us point out that this problem becomes undecidable for strict,rational semi-Thue systems ([O'D81, O'D83])(hence, a fortiori for strict, finite c-systems).

**Theorem 4.5**
(1) *Given $S_1, S_2$ strict,rational,confluent semi-Thue systems, one can decide whether* $\xleftrightarrow[S_1]{*} = \xleftrightarrow[S_2]{*}$
(2) *Given $S_1, S_2$ strict,finite,confluent,left-basic c-systems, one can decide whether* $\xleftrightarrow[S_1]{*} \subseteq \xleftrightarrow[S_2]{*}$

Point (1) is shown in [Nar86] and point (2) in [Sén90b].Let us point out that problem (1) becomes undecidable for strict,finite,confluent c-systems ([Sén90b]).

**Theorem 4.6**
(1) *The property "$S$ is partly confluent on $R$"* [11] *where $S$ is some strict, finite, basic semi-Thue system and $R \subseteq Irr(S)$ is rational, is decidable*
(2) *The property "$S$ is confluent on $\Delta_S^*(R)$" where $S$ is a strict, finite, monadic semi-Thue system and $R$ is a rational set, is decidable.*

Point (1) is shown in [Ott87] in the monadic case and for $R$ reduced to a single word, and in [Sén90b] in the above case. The proofs are similar and lean essentially on Valiant's theorem ([Val74, Bee76], see footnote 10). Point (2) is shown in [Nar90]. Let us notice that (1) (resp.(2))becomes undecidable for semi-Thue systems $S$ which are assumed strict, finite [Ott87, Sén90b] (resp. for semi-Thue systems $S$ which are assumed l-decreasing, noetherian, finite [Car91]). The property to be "partly confluent" (as well as other analogous properties)is studied in more details in [OZ91, Ott92a, Ott92b, MNOZ93, Zha92].

[11] see the definition given in §2.1

# 5 CONTEXT-FREE LANGUAGES & WORD-REWRITING

We do not know any theorem characterising context-free languages in terms of word rewriting systems (in the sense defined in the introduction).One can nevertheless describe these languages by means of word rewriting systems and of *morphisms* of free-monoïds: it suffices to notice that context-free languages are the homomorphic images of deterministic context-free languages, and then to apply theorem 4.1. A tighter statement is the following

**Theorem 5.1 ([But73b, Ber77])** *Let $L \subseteq X^+$. The following properties are equivalent*

**(1)** *$L$ is context-free*

**(2)** *$\exists Y \supseteq X, \exists F \subseteq Y^*, F$ finite,$\exists S$ strict,finite,basic,confluent semi-Thue system,$\exists \varphi : Y^* \longrightarrow X^*$, strictly alphabetic morphism, such that*

$$L = \varphi([F]_{\overset{*}{\underset{S}{\longleftrightarrow}}})$$

Sketch of proof:

(2) $\Longrightarrow$ (1): follows from theorem 4.1

(1) $\Longrightarrow$ (2): $L$ is generated by some context-free grammar in Greibach normal form. Hence $L$ is the image by some strictly alphabetic morphism of some "very simple" language $M$. But every very simple language has the form $[F]_{\overset{*}{\underset{S}{\longleftrightarrow}}}$ where $F, S$ are fulfilling the conditions of property (2) ([But73b]).

The family of all the context-free languages of the form $[f]_{\overset{*}{\underset{S}{\longleftrightarrow}}}$ with $S$ strict , finite, confluent semi-Thue system, is a sub-family of the family of context-free languages which has no known characterisation in terms of grammars. A study of this family is presented in [Coc76].

Here are some examples of languages in this sub-family:

**(1)** Point (3) of theorem 4.1, enables one to generate a *deterministic* context-free example from every given D.P.D.A.

**(2)** The miror images of the examples of type (1) give examples which are *deterministic-miror* (but not necessarily deterministic)

**(3)** The union of some (suitably choosen) example of type (1) with some (suitably choosen) example of type (2) constructed on disjoint alphabets, gives an example which is *neither deterministic, nor deterministic-miror*, but remains *non-ambiguous.*

**(4)** It is possible to build ambiguous examples (i.e. *inherently ambiguous*). Such an example is given in [Sén87, p.301-308].This example leans on a theorem of [Ber86] and a theorem of [AFG86]; it consists essentially in "encoding" the Thue-Morse homomorphism ([Thu12, MH44]) into a strict,finite, confluent semi-Thue system.

Very few is known about the set of all pairs $(f, S)$ such that $[f]_{\overset{*}{\underset{S}{\longleftrightarrow}}}$ is context-free ( excepted the results exposed in part 4). The following theorem shows that it is not recursive.

**Theorem 5.2** ([NOR86]) *Let $\Omega$ be some family of context-free languages which contains all the finite languages. Then*

**(1)** *$\exists S$ strict,finite,confluent semi-Thue system such that the question, given $w \in X^*$*

$$[w]_{\overset{*}{\underset{S}{\longleftrightarrow}}} \in \Omega?$$

*is undecidable (even $\Sigma_1$-hard)*

**(2)** *The question , given some finite semi-Thue system $S$*

$$\forall w \in X^*, [w]_{\overset{*}{\underset{S}{\longleftrightarrow}}} \in \Omega?$$

*is undecidable (even $\Pi_2$-hard)*

Let us notice that theorem 5.2 applies on the particular families:

$$\Omega_1 = \text{family of all rational languages}$$

and

$$\Omega_2 = \text{family of all deterministic context-free languages}$$

though these families can be described in terms of some families of finite semi-Thue systems or c-systems (see sections 3 et 4).
The statement obtained by choosing $\Omega = \Omega_1$ in theorem 5.2 solves negatively the question raised in [Coc71, p.72].

# 6 RECURSIVELY ENUMERABLE LANGUAGES & WORD-REWRITING

**Theorem 6.1 ([O'D81, Ott84])** *Let $L \subset X^*$. The following properties are equivalent*

**(1)** *$L$ is recursively enumerable*
**(2)** *$\exists Y \supseteq X \cup \{\#, \bar{\#}, \$\}$, $\exists S$ strict,finite,confluent semi-Thue system and $\exists R \subseteq Y^*$ rational such that*

$$\#\bar{\#}L\$ = [R]_{\underset{S}{\overset{*}{\longleftrightarrow}}} \cap \#\bar{\#}X^*\$ = \Delta_S^*(R) \cap \#\bar{\#}X^*\$$$

**Theorem 6.2 ([BJW82])** *Let $L \subset X^*$. The following properties are equivalent*

**(1)** *$L$ is recursively enumerable*
**(2)** *$\exists Y \supseteq X$, $\exists S$ finite,special,confluent semi-Thue system and $\exists C \subseteq Y^*$ context-free such that*

$$L = [C]_{\underset{S}{\overset{*}{\longleftrightarrow}}} \cap X^* = \Delta_S^*(C) \cap X^*$$

These two theorems are somewhat "antithetic" to " Benois' theorem" (theorem 3.5). These theorems enable one to give "elegant" proofs (i.e. which do not consist in direct encodings of Turing machines or of instances of Post correspondence problem) of a great number of undecidability results about semi-Thue systems ([Ott84, Ott87, Sén90b]).

Let us give one application to formal language theory:

**Application 6.3 ([Sén90b])** *The following problem is undecidable*

*instance: Some word $f$ and two semi-Thue systems $S_1, S_2$ assumed finite,strict and confluent.*

*question :* $[f]_{\underset{S_1}{\overset{*}{\longleftrightarrow}}} = [f]_{\underset{S_2}{\overset{*}{\longleftrightarrow}}}$ *?*

This result is a negative answer to the question raised in [NB72, p.9]. Let us notice that the undecidability of this problem stems from the fact that here $f$ is given in the instance (compare with theorem 4.5). Let us notice further that:

- if, in addition , it is assumed that $S_1$ or $S_2$ is basic, then the problem becomes decidable ([Sén89])
- if, in addition it is assumed that $S_1$ or $S_2$ is left-basic, then the problem becomes equivalent to the general equivalence problem for D.P.D.A. ([Sén90b],see section 4).

# 7 Acknowledgements

I thank L. Boasson for his comments on a first version of the paper.

## References

[AB92] J.M. Autebert and L. Boasson. The equivalence of pre-NTS grammars is decidable. *Math.Systems Theory 25*, pages 61–74, 1992.

[AFG86] J.M. Autebert, P. Flajolet, and J. Gabarro. Prefixes of infinite words and ambiguous context-free languages. *IPL 25 (4)*, 1986.

[Aut87] J.M. Autebert. *Langages algébriques*. Masson, 1987.

[BB91] J. Berstel and L. Boasson. Context-free languages. In *Handbook of theoretical computer science, vol.B, Chapter 2*, pages 59–102. Elsevier, 1991.

[Bee76] C. Beeri. An improvement on Valiant's decision procedure for equivalence of deterministic finite-turn pushdown automata. *TCS 3*, pages 305–320, 1976.

[Ben69] M. Benois. Parties rationnelles du groupe libre. *C.R.Acad.Sci.série A*, pages 1188–1190, 1969.

[Ben87] M. Benois. Descendants of regular language in a class of rewriting systems:algorithm and complexity of an automata construction. *Proceedings RTA 87, LNCS 256*, pages 121–132, 1987.

[Ber77] J. Berstel. Congruences plus que parfaites et langages algébriques. *Séminaire d'informatique théorique de Paris6-7*, 1976-1977.

[Ber86] J. Berstel. Every iterated morphism yields a co-cfl. *IPL 22*, pages 7–9, 1986.

[BJW82] R.V. Book, M. Jantzen, and C. Wrathall. Monadic Thue systems. *TCS 19*, pages 231–251, 1982.

[BN84] L. Boasson and M. Nivat. Centers of context-free languages. *LITP technical report no84-44*, 1984.

[BO85] R.V. Book and F. Otto. Cancellation rules and extended word problems. *IPL 20*, pages 5–11, 1985.

[BO93] R.V. Book and F. Otto. *String Rewriting Systems.Texts and monographs in Computer Science*. Springer-Verlag, 1993.

[Boa80] L. Boasson. Grammaires à non-terminaux séparés. In *Proceedings 7th ICALP*, pages 105–118. LNCS 85, 1980.

[Boo87] R.V. Book. Thue systems as rewriting systems. *J.Symbolic Computation.3*, pages 39–68, 1987.

[BS86] M. Benois and J. Sakarovitch. On the complexity of some extended word problems defined by cancellation rules. *IPL 23*, pages 281–287, 1986.

[Büc64] J.R. Büchi. Regular canonical systems. *Archiv fur Math.Logik und Grundlagenforschung 6*, pages 91–111, 1964.

[Büc90] J.R. Büchi. Regular canonical systems. *the collected works of J.R. Buchi*, pages 317–337, 1990.

[But73a] P. Butzbach. Sur l'équivalence des grammaires simples. In *Actes des Premières journées d'Informatique Théorique,Bonascre*. ENSTA, 1973.

[But73b] P. Butzbach. Une famille de congruences de Thue pour lesquelles l'équivalence est décidable. In *Proceedings 1rst ICALP*, pages 3–12. LNCS, 1973.

[Car91] A.C. Caron. Linear bounded automata and rewrite systems : influence of initial configurations on decision properties. In *Proceedings CAAP 91*, pages 74–89. LNCS ,Springer-Verlag, 1991.

[Cau88] D. Caucal. Recritures suffixes de mots. *Internal report IRISA nr 413*, 1988.

[Cau89] D. Caucal. A fast algorithm to decide on simple grammars equivalence. *L.N.C.S., Vol.401*, pages 66–85, 1989.

[Cau93] D. Caucal. A fast algorithm to decide on the equivalence of stateless dpda. *RAIRO Theoret.Infor.and App.*, pages 23–48, 1993.

[Cho79] L. Chottin. Strict deterministic languages and controlled rewriting systems. In *Proceedings 6th ICALP*. LNCS Springer-Verlag, 1979.

[Cho82] L. Chottin. Langages algébriques et systèmes de réécriture rationnels. *RAIRO Inform.Téor.16*, pages 1–20, 1982.

[CN71] Y. Cochet and M. Nivat. Une généralisation des ensembles de Dyck. *Israel J. of Math. vol.9, n 3*, 1971.

[Coc71] Y. Cochet. Sur l'algébricité des classes de certaines congruences définies sur le monoide libre. *Thèse de l'université de Rennes*, 1971.

[Coc75] Y. Cochet. Congruences quasi-parfaites et langages rationnels. *Manuscrit non-publié*, 1975.

[Coc76] Y. Cochet. Langages définis par des congruences. *Groupe d'étude d'Algèbre ,1ère année,no 16*, pages 1–7, 1975/76.

[DJ91] N. Dershowitz and J.P. Jouannaud. Rewrite systems. In *Handbook of theoretical computer science, vol.B, Chapter 2*, pages 243–320. Elsevier, 1991.

[EHR83] A. Ehrenfeucht, D. Haussler, and G. Rozenberg. On regularity of context-free languages. *T.C.S.27*, pages 311–322, 1983.

[GG66] S. Ginsburg and S. Greibach. Deterministic context-free languages. *Information and Control 9*, pages 620–648, 1966.

[Gre67] S. Greibach. A note on pushdown store automaton and regular systems. *Proc.Amer.Math.Soc.*, pages 263–268, 1967.

[Har78] M.A. Harrison. *Introduction To Formal Language Theory*. Addison-Wesley, 1978.

[HU79] J.E. Hopcroft and J.D. Ullman. *Introduction to Automata Theory,Languages, and Computation*. Addison-Wesley, 1979.

[Hue80] G. Huet. Confluent reductions:abstract properties and applications to term rewriting systems. *JACM vol. 27 no 4*, pages 797–821, 1980.

[Jan88] M. Jantzen. *Confluent String Rewriting (EATCS monograph)*. Springer-Verlag, 1988.

[Kie79] F. Kierszenbaum. Les langages à opérateurs d'insertion. *Thèse de 3ième cycle , université Bordeaux I*, 1979.

[KS86] W. Kuich and A. Salomaa. *Semirings,Automata,Languages (EATCS monograph)*. Springer-Verlag, 1986.

[Lal94] G. Lallement. The word problem for Thue systems. In *Actes de l'école de printemps d'informatique théorique, Font-Romeu 93*. Springer-Verlag, 1994.

[LP94] Y. Lafont and A. Prouté. Church-Rosser property and homology of monoids. In *Actes de l'école de printemps d'informatique théorique, Font-Romeu 93*. Springer-Verlag, 1994.

[Mat94] Y. Matiyasevich. Word problem for Thue systems with a few relations. In *Actes de l'école de printemps d'informatique théorique, Font-Romeu 93*. Springer-Verlag, 1994.

[McC93] R. McCloskey. Bounded context grammars and languages. *PHD*, 1993.

[McN67] R. McNaughton. Parenthesis grammars. *JACM 14*, pages 490–500, 1967.

[MH44] M. Morse and G.A. Hedlung. Unending chess,symbolic dynamics and a problem in semi-groups. *Duke Math. J.11*, pages 1–7, 1944.

[MNO88] R. McNaughton, P. Narendran, and F. Otto. Church-Rosser Thue systems and formal languages. *JACM 35 (2)*, pages 324–344, 1988.

[MNOZ93] K. Madlener, P. Narendran, F. Otto, and L. Zhang. On weakly confluent monadic string-rewriting systems. *T.C.S. 113*, pages 119–165, 1993.

[MO89] K. Madlener and F. Otto. About the descriptive power of certain classes of finite string-rewriting systems. *T.C.S. 67*, pages 143–172, 1989.

[Nar86] P. Narendran. On the equivalence problem for regular Thue systems. *TCS 44*, pages 237–245, 1986.

[Nar90] P. Narendran. It is decidable whether a monadic Thue system is canonical over a regular set. *Math. Systems Theory 23*, pages 245–254, 1990.

[NB72] M. Nivat and M. Benois. Congruences parfaites et quasi-parfaites. *Séminaire Dubreil,2ième année,no 7*, 1971/72.

[Niv70] M. Nivat. On some families of languages related to the Dyck language. *2nd Annual Symposium on Theory of Computing*, 1970.

[Niv71] M. Nivat. Congruences de Thue et t-langages. *Studia Scientiarum Mathematicarum Hungarica 6*, pages 243–249, 1971.

[NOR86] P. Narendran, C. O'Dunlaing, and H. Rolletschek. Complexity of certain decision problems about congruential languages. *JCSS 30*, pages 343–358, 1986.

[O'D81] C. O'Dunlaing. Finite and infinite regular Thue systems. *PHD University of California at Santa Barbara*, 1981.

[O'D83] C. O'Dunlaing. Infinite regular thue systems. *TCS 2*, pages 171–192, 1983.

[OIH81] M. Oyamaguchi, Y. Inagaki, and N. Honda. The equivalence problem for two dpda's one of which is a finite-turn or one-counter machine. *J. Comput. System Sci.23,nr 3*, pages 366–382, 1981.

[Ott84] F. Otto. Some undecidability results for non-monadic Church-Rosser Thue systems. *TCS 33*, pages 261–278, 1984.

[Ott87] F. Otto. On deciding the confluence of a finite string-rewriting system on a given congruence class. *JCSS 35*, pages 285–310, 1987.

[Ott92a] F. Otto. Completing a finite special string-rewriting system on the congruence class of the empty word. *Applicable Algebra in Engeneering Comm.Comput. 2*, pages 257–274, 1992.

[Ott92b] F. Otto. The problem of deciding confluence on a given congruence class is tractable for finite special string-rewriting systems. *Math. Systems Theory 25*, pages 241–251, 1992.

[OZ91] F. Otto and L. Zhang. Decision problems for finite special string-rewriting systems that are confluent on some congruence class. *Acta Informatica 28*, pages 477–510, 1991.

[RS86] F. Romian and J. Sakarovitch. One-sided Dyck reduction over two letters and deterministic context-free languages. In *Proceedings MFCS 86*, pages 554–563. Springer-Verlag, LNCS 233, 1986.

[Sak79] J. Sakarovitch. Syntaxe des langages de Chomsky, essai sur le déterminisme. *Thèse de doctorat d'état de l'université Paris VII*, pages 1–175, 1979.

[Sal73] A. Salomaa. *Formal Languages*. Academic Press, 1973.

[Sén81] G. Sénizergues. Décidabilité de l'équivalence des grammaires NTS. *Thèse de 3ième cycle,université Paris 7*, pages 1–170, 1981.

[Sén85] G. Sénizergues. The equivalence and inclusion problems for NTS languages. *J. Comput. System Sci. 31(3)*, pages 303–331, 1985.

[Sén86] G. Sénizergues. On the rational subsets of the free group. *LITP internal report nr 86-3, submitted to Acta Informatica*, pages 1–17, 1986.

[Sén87] G. Sénizergues. Sur la description des langages algébriques déterministes par des systèmes de réécriture confluents. *Thèse d'état,université Paris 7 et rapport LITP 88-39*, pages 1–330, 1987.

[Sén89] G. Sénizergues. Church-Rosser controlled rewriting systems and equivalence problems for deterministic context-free languages. *Information and Computation 81 (3)*, pages 265–279, 1989.

[Sén90a] G. Sénizergues. A characterisation of deterministic context-free languages by means of right-congruences. *TCS 70 (2)*, pages 213–232, 1990.

[Sén90b] G. Sénizergues. Some decision problems about controlled rewriting systems. *TCS 71*, pages 281–346, 1990.

[Tak75] M. Takahashi. Generalizations of regular sets and their application to a study of context-free languages. *Information and Control 27*, pages 1–35, 1975.

[Thu12] A. Thue. Uber die gegenseitiges lage gleicher teile gewisser zeichenreihen. *Skr. Vid.Kristiania,I Mat.Naturv.Klasse nr 8*, pages 1–67, 1912.

[Thu14] A. Thue. Probleme über veränderungen von zeichenreihen nach gegeben regeln. *Skr. Vid.Kristiania,I Mat.Naturv.Klasse nr 10,*, page 34 pp, 1914.

[Val74] L.G. Valiant. The equivalence problem for deterministic finite-turn pushdown automata. *Information and Control 25*, pages 123–133, 1974.

[Zha92] L. Zhang. The pre-NTS property is undecidable for context-free grammars. *IPL 44*, pages 181–184, 1992.