Locally Stratified Boolean Grammars *

Christos Nomikos[1] and Panos Rondogiannis[2]

[1] Department of Computer Science, University of Ioannina,
P.O. Box 1186, 45 110 Ioannina, Greece
cnomikos@cs.uoi.gr

[2] Department of Informatics & Telecommunications
University of Athens, Athens, Greece
prondo@di.uoa.gr

Abstract. We introduce *locally stratified boolean grammars* which form a natural subclass of boolean grammars [*A. Okhotin, Information and Computation, 194 (2004) 19-48*] with many desirable properties. This new class of grammars extends the stratified ones [*M. Wrona, MFCS (2005) 801-812*]. We demonstrate that the property of local stratifiability can be tested in linear time with respect to the size of the grammar under consideration. We then develop the semantics of locally stratified grammars and demonstrate that it coincides with their well-founded semantics [*V. Kountouriotis et al., DLT (2006) 203-214*]; moreover, we show that for each such grammar the well-founded semantics is total (ie., two-valued). Finally, we use the new semantics in order to show that locally stratified grammars can even express certain languages that are currently not known to be expressible by the stratified semantics (such as for example the language $\{a^{2^n} \mid n \geq 0\}$).

1 Introduction

Boolean grammars were recently proposed by A. Okhotin [Okh04] as a means to overcome certain expressibility limitations of context-free grammars. Intuitively, boolean grammars allow intersection and complementation to appear in the right hand sides of rules; these two operations combined with the inherent recursion of context-free rules, make this new formalism very powerful (but still tractable from a parsing point of view).

As it became obvious from the very beginning, the problem of assigning a proper semantics to boolean grammars is non-trivial. The difficulties arise due to the fact that the rules of such grammars may contain circularities that "pass through negation" and that are not easy to handle. Initially, two semantics were proposed [Okh04], namely the *unique solution* and the *naturally feasible*

* This research was co-funded by the European Union in the framework of the project "Support of Computer Science Studies in the University of Ioannina" of the "Operational Program for Education and Initial Vocational Training" of the 3rd Community Support Framework of the Hellenic Ministry of Education, funded by 25% from national sources and by 75% from the European Social Fund (ESF).

ones; however, both of these approaches exhibit some undesirable behavior in certain cases (see [KNR06] for more details). More recently, based on well-known ideas from logic programming [vGRS91], Kountouriotis et al. proposed the *well-founded semantics* of boolean grammars [KNR06]. This latter approach requires the study of *three-valued languages*, i.e, languages in which the membership of a string can be classified as true, false or unknown. The advantage of this approach is that it applies to all boolean grammars, independently of their syntax.

Another direction of research which was initiated by M. Wrona [Wro05] seeks to find subclasses of boolean grammars that are well-behaved both semantically and from an application point of view. More specifically, in [Wro05] the class of *stratified boolean grammars* is defined and their properties are investigated. The notion of stratification also has its roots in the area of logic programming (see for example [ABW88]).

In this paper, motivated again by ideas in logic programming [Prz88], we propose the *locally stratified boolean grammars* which form a proper superset of the stratified ones. We demonstrate that the property of local stratifiability can be tested in linear time with respect to the size of the grammar under consideration. This is a surprising fact because local stratifiability in logic programming is undecidable (more specifically Π^1_1-complete [CB94]). We then develop the semantics of locally stratified grammars and demonstrate that it coincides with their well-founded semantics [KNR06]; moreover, we show that for each such grammar the well-founded semantics is total (ie., two-valued). Finally, we use the new semantics in order to show that locally stratified grammars can even express certain languages that are currently not known to be expressible by the stratified semantics (such as for example the language $\{a^{2^n} \mid n \geq 0\}$).

2 Boolean Grammars

In [Okh01] and [Okh04] A. Okhotin introduced the classes of conjunctive and boolean grammars respectively. Formally:

Definition 1 ([Okh04]). *A* boolean grammar *is a quadruple $G = (\Sigma, N, P, S)$, where Σ and N are disjoint finite nonempty sets of terminal and nonterminal symbols respectively, P is a finite set of rules, each of the form*

$$C \to \alpha_1 \& \cdots \& \alpha_m \& \neg\beta_1 \& \cdots \& \neg\beta_n \qquad (m + n \geq 1, C \in N, \alpha_i, \beta_i \in (\Sigma \cup N)^*)$$

and $S \in N$ is the start symbol of the grammar. We will call the α_i's positive literals *and the $\neg\beta_i$'s* negative*. A boolean grammar is called* conjunctive *if all its rules contain only positive literals.*

In this paper we introduce a large subclass of boolean grammars which has a well-defined semantics. We will need the following definitions:

Definition 2. *An* interpretation *I of a boolean grammar $G = (\Sigma, N, P, S)$ is a function $I : N \to 2^{\Sigma^*}$.*

Notice that the above notion of interpretation is classical and not three-valued as in the case of the well-founded semantics (see [KNR06]). In simpler terms, the locally stratified boolean grammars that we will define shortly, do not require the generality of the semantic machinery needed for general boolean grammars.

In the following we denote by $\perp$ the interpretation which assigns to every non-terminal symbol of a grammar the empty set. An interpretation I can be recursively extended to apply to expressions that appear as the right-hand sides of boolean grammar rules:

Definition 3. *Let $G = (\Sigma, N, P, S)$ be a boolean grammar and I be an interpretation of G. Then I can be extended to become a* truth valuation *$\hat{I}$ as follows:*

- $\hat{I}(\epsilon) = \{\epsilon\}$.
- $\hat{I}(a) = \{a\}$, *for every $a \in \Sigma$.*
- $\hat{I}(A) = I(A)$, *for every $A \in N$.*
- $\hat{I}(x_1 x_2 \dots x_n) = \hat{I}(x_1) \circ \hat{I}(x_2) \circ \cdots \circ \hat{I}(x_n)$, *for every sequence $x_1 x_2 \dots x_n \in (\Sigma \cup N)^*$ (where $\circ$ is the usual concatenation operator for languages).*
- $\hat{I}(\neg\alpha) = \Sigma^* - \hat{I}(\alpha)$, *for every $\alpha \in (\Sigma \cup N)^*$.*
- $\hat{I}(l_1 \& l_2 \& \dots \& l_n) = \hat{I}(l_1) \cap \hat{I}(l_2) \cap \cdots \cap \hat{I}(l_n)$, *for every sequence $l_1, l_2, \dots, l_n$ of literals.*

The locally stratified boolean grammars form a proper superset of the class of stratified grammars introduced by M. Wrona. In the following definiton (as-well-as elsewhere in the rest of the paper), we denote by ω the set of natural numbers.

Definition 4 ([Wro05]). *A boolean grammar $G = (\Sigma, N, P, S)$ is called* stratified *if there exists a function $g : N \to \omega$ such that for every rule*

$$C \to \alpha_1 \& \cdots \& \alpha_m \& \neg\beta_1 \& \cdots \& \neg\beta_n$$

in P the following conditions hold:

- *for every i, $1 \leq i \leq m$ and for every $A \in N$ that appears in α_i, $g(C) \geq g(A)$*
- *for every j, $1 \leq j \leq n$ and for every $B \in N$ that appears in β_j, $g(C) > g(B)$.*

The class of stratified boolean grammars defined above is obviously a proper subclass of boolean grammars, but it appears to have an interest in its own right. For example, questions of the form "are there languages that can be defined by general boolean grammars but not from stratified ones?" do not in general have obvious answers (and may trigger deeper investigations in the theory of these grammars).

3 Locally Stratified Boolean Grammars

In this section we introduce a class of boolean grammars that is broader than the stratified ones and that can express certain languages that are currently not known to be expressible by stratified grammars. To motivate the new class, consider the following boolean grammar with start symbol $S = Even$, that defines the (regular) set of strings of even length over the alphabet $\Sigma = \{a\}$:

$$
\begin{aligned}
&Even \rightarrow \epsilon \\
&Even \rightarrow a\ Odd \\
&Odd\ \ \rightarrow \neg Even
\end{aligned}
$$

One can easily see that the above grammar is not stratified. However, it can easily be seen that the grammar specifies the language we mentioned above. For example, the string aa belongs to the language corresponding to *Even* because the string a belongs to the language corresponding to *Odd* (since it does not belong to the language corresponding to *Even*).

Grammars such as the above are locally stratified. Informally, if a grammar is locally stratified then the pairs in $(N \times \Sigma^*)$ can be partitioned into a (possibly infinite) set of strata so that if the membership of w in the language defined by nonterminal C depends on the membership of w' in the language defined by nonterminal D, then (D, w') cannot belong to a stratum higher than the stratum of (C, w); furthermore if the above dependency is obtained through negation, (D, w') must belong to a stratum lower than the stratum of (C, w). Formally:

Definition 5. *A boolean grammar $G = (\Sigma, N, P, S)$ is locally stratified if there exists a function $f : (N \times \Sigma^*) \rightarrow \omega$ such that for every rule*

$$C \rightarrow \alpha_1 \& \cdots \& \alpha_m \& \neg\beta_1 \& \cdots \& \neg\beta_n$$

in P, the following conditions hold for every i, $1 \leq i \leq m$ and for every j, $1 \leq j \leq n$:

- *Suppose that $\alpha_i = \sigma_1 A_1 \sigma_2 A_2 \ldots \sigma_k A_k \sigma_{k+1}$, for $k \geq 1, \sigma_p \in \Sigma^*, A_p \in N$. Then for every $w_1, w_2, \ldots, w_k \in \Sigma^*$ and for every p, $1 \leq p \leq k$, it holds $f(C, \sigma_1 w_1 \sigma_2 w_2 \ldots \sigma_k w_k \sigma_{k+1}) \geq f(A_p, w_p)$.*
- *Suppose that $\beta_j = \tau_1 B_1 \tau_2 B_2 \ldots \tau_\ell B_\ell \tau_{\ell+1}$, for $\ell \geq 1, \tau_q \in \Sigma^*, B_q \in N$. Then for every $w_1, w_2, \ldots, w_\ell \in \Sigma^*$ and for every q, $1 \leq q \leq \ell$, it holds $f(C, \tau_1 w_1 \tau_2 w_2 \ldots \tau_\ell w_\ell \tau_{\ell+1}) > f(B_q, w_q)$.*

As we have already mentioned, local stratification is a notion that originates from logic programming [Prz88]. There are however some crucial differences that make the study of local stratification in boolean grammars even more interesting. First, local stratification of logic programs in many cases requires a transfinite number of strata (ie., the labeling of the strata can not in general be performed solely by members of ω but may additionally require the use of ordinal numbers that are greater than ω). Second, the problem of detecting whether a logic program is locally stratified is in general unsolvable (see [CB94]) and one can only hope to find subclasses of logic programs in which the notion is decidable (see for example [Ron01,NRG05]). Surprisingly, it turns out that local stratifiability of boolean grammars can be decided in polynomial time.

In order to define the semantics of locally stratified boolean grammars (see next section), it is convenient to use a stratum function that has some special properties.

Definition 6. *Let $G = (\Sigma, N, P, S)$ be a boolean grammar locally stratified by a function f. We say that f is a canonical stratum-function if*

- *for every $w, w' \in \Sigma^*$ and for every $A, B \in N$, if $|w| > |w'|$ then $f(A, w) > f(B, w')$.*
- *for every $w, w' \in \Sigma^*$ and for every $A \in N$, if $|w| = |w'|$ then $f(A, w) = f(A, w')$.*

We can now demonstrate that local stratifiability of boolean grammars is decidable (and actually, efficiently so). Before we state Theorem 1 that proves this fact, we need the following definition:

Definition 7. *Let $G = (\Sigma, N, P, S)$ be a boolean grammar. The skeleton of G is the grammar $G' = (\Sigma, N, P', S)$, where P' is obtained from P by removing from the right-hand side of each rule every literal that equals ϵ or $\neg\epsilon$, or contains terminal symbols and then removing all rules that end up with an empty right-hand side.*

Theorem 1. *A boolean grammar $G = (\Sigma, N, P, S)$ is locally stratified if and only if its skeleton $G' = (\Sigma, N, P', S)$ is stratified.*

Proof. Suppose that G is locally stratified by f. Define a function $g : N \to \omega$ such that $g(A) = f(A, \epsilon)$. Let $C \to \alpha_1 \& \cdots \& \alpha_m \& \neg\beta_1 \& \cdots \& \neg\beta_n$ be a rule in P'. Suppose that $A \in N$ appears in some α_i. Since G' is the skeleton of G, α_i is of the form $A_1 A_2 \ldots A_k$, $k \geq 1$ and $A_p \in N$ for $1 \leq p \leq k$, and $A = A_r$ for some r, $1 \leq r \leq k$. Notice that, $\alpha_i = \sigma_1 A_1 \sigma_2 A_2 \ldots \sigma_k A_k \sigma_{k+1}$, where $\sigma_1 = \sigma_2 = \ldots = \sigma_k = \sigma_{k+1} = \epsilon$. Let $w_1 = w_2 = \ldots = w_k = \epsilon$. By the definition of local stratification we get $f(C, \epsilon) \geq f(A_r, \epsilon)$, which implies $g(C) \geq g(A_r) = g(A)$. Similarly it can be proved that if $B \in N$ appears in some β_j, then $g(C) > g(B)$. Consequently, G' is stratified by g.

Conversely, suppose that the skeleton G' is stratified by g and let $s = 1 + \max\{i \in \omega \mid \exists A \in N \text{ such that } g(A) = i\}$. In other words s is an upper bound for the number of the non-empty strata according to g. Define $f : (N \times \Sigma^*) \to \omega$ such that $f(A, w) = s \cdot |w| + g(A)$. It is easy to see that f is a canonical stratum-function. Let $C \to \alpha_1 \& \cdots \& \alpha_m \& \neg\beta_1 \& \cdots \& \neg\beta_n$ be a rule in P. Consider an $\alpha_i = \sigma_1 A_1 \sigma_2 A_2 \ldots \sigma_k A_k \sigma_{k+1}$ and an arbitrary sequence of strings $w_1, w_2, \ldots, w_k \in \Sigma^*$. Let $w = \sigma_1 w_1 \sigma_2 w_2, \ldots \sigma_k w_k \sigma_{k+1}$. If $|w| > |w_p|$, then $f(C, w) > f(A_p, w_p)$ by the canonicity of f. Otherwise ($|w| = |w_p|$) it holds $\sigma_1 = \sigma_2 = \cdots = \sigma_k = \sigma_{k+1} = \epsilon$, i.e $\alpha \in N^*$. Therefore G' contains a rule with C in the left-hand side and α_i in the right-hand side, which implies $g(C) \geq g(A_p)$. Since $|w| = |w_p|$, by the definition of f we get $f(C, w) \geq f(A_p, w_p)$. The case for β_j is similar. Consequently, G is locally stratified by f. $\square$

Corollary 1. *A boolean grammar G is locally stratified if and only if it is locally stratified by a canonical stratum-function.*

Corollary 2. *If a boolean grammar G is stratified then it is locally stratified.*

Proof. If G is stratified by f then its skeleton is also stratified by f. □

The converse of Corollary 2 does not hold as the example in the beginning of this section as well as the following one demonstrate:

Example 1. Consider the boolean grammar $G = (\Sigma, N, P, S)$, where $\Sigma = \{a\}$ and P contains the following rules:

$$\begin{array}{llll} S \to X \& \neg aX & A \to \neg X & D \to \neg Y \& T & Z \to Y \\ S \to \neg X \& aX & B \to \neg AA & E \to \neg DD \& T & Z \to aY \\ S \to Z \& \neg aZ & C \to \neg BB & F \to \neg EE \& T & T \to \epsilon \\ S \to \neg Z \& aZ & X \to aCC & Y \to aaFF & T \to aaT \end{array}$$

The above grammar defines the language $L = \{a^{2^n} \mid n \in \omega\}$ (see [Okh04]). This grammar is not stratified; furthermore it is not known if L can be expressed by a stratified grammar [Wro05]. However, G is locally stratified. To prove this claim, consider the skeleton G' of G which contains the following set of rules:

$$\begin{array}{llllll} S \to X & S \to Z & A \to \neg X & C \to \neg BB & E \to \neg DD \& T & Z \to Y \\ S \to \neg X & S \to \neg Z & B \to \neg AA & D \to \neg Y \& T & F \to \neg EE \& T & \end{array}$$

It is easy to see that G' is stratified by the function g, such that $g(X) = g(Y) = g(Z) = g(T) = 0$, $g(A) = g(D) = g(S) = 1$, $g(B) = g(E) = 2$ and $g(C) = g(F) = 3$. Thus G is locally stratified.

The above theorem shows that testing local stratifiability of a boolean grammar G can be reduced to testing (ordinary) stratifiability of the skeleton of G. The reduction requires time $\mathcal{O}(|G|)$, where $|G|$ denotes the size of the representation of G, and produces a grammar G' with $|G'| \leq |G|$. Testing if G' is stratified requires time $\mathcal{O}(|G'|)$ [Wro05], using simple graph algorithms. Consequently local stratifiability of a boolean grammar G can be tested in time $\mathcal{O}(|G|)$. Notice that, as we have already mentioned, testing local stratifiability of logic programs is an unsolvable problem.

4 Locally Stratified Semantics for Boolean Grammars

In this section we demonstrate how one can define the semantics of a boolean grammar that is locally stratified. The languages defined by the non-terminal symbols in a locally stratified boolean grammar, can be constructed in stages. During the i-th stage, for every pair (A, w) that belongs to the i-th stratum we decide whether w belongs to the language defined by A. The following definition will be needed:

Definition 8. *Let Σ be an alphabet. We denote by Σ^n the set $\{w \in \Sigma^* \mid |w| = n\}$ and by $\Sigma^{\leq n}$ the set $\bigcup_{i=0}^{n} \Sigma^i$.*

Let $G = (\Sigma, N, P, S)$ be a boolean grammar, I an interpretation, $M \subseteq N$ be a set of non-terminal symbols, and $n \geq 0$ be an integer. We first define the conjunctive grammar $G/(I, M, n)$ that is used to decide the membership of strings of length n, in the languages corresponding to symbols in M (as defined by the rules in G), provided that some subsets of these languages are known and determined by I. Formally:

Definition 9. *Let $G = (\Sigma, N, P, S)$ be a Boolean grammar, I be an interpretation, $M \subseteq N$ be a set of non-terminal symbols, and $n \geq 0$ be an integer. Let R be the set of all literals that appear in the right hand sides of the rules in P in which the left-hand side symbol is in M. We denote by $G/(I, M, n)$ the grammar (Σ, N', P', S), such that:*

- *$N' = N \cup \{D_l \mid l \in R\}$, where the D_l's are new non-terminal symbols not belonging to N.*
- *For every rule of the form $C \to l_1 \& l_2 \& \ldots \& l_m$ in P, such that $C \in M$, P' contains the rule: $C \to D_{l_1} \& D_{l_2} \& \ldots \& D_{l_m}$.*
- *For every literal $l \in R$ and for every $w \in (\hat{I}(l) \cap \Sigma^n)$, P' contains the rule $D_l \to w$.*
- *if $n > 0$ then for every literal $l = A_1 A_2 \cdots A_k \in R \cap N^+$ and for every i, $1 \leq i \leq k$, if $\epsilon \in \bigcap_{1 \leq j \leq k, j \neq i} \hat{I}(A_j)$ then P' contains the rule $D_l \to A_i$.*
- *if $n = 0$ then for every literal $l = A_1 A_2 \cdots A_k \in R \cap N^*$, P' contains the rule $D_l \to l'$, where $l' = \alpha_1 \alpha_2 \cdots \alpha_k$, with $\alpha_i = \epsilon$ if $\epsilon \in I(A_i)$ and $\alpha_i = A_i$ otherwise.*

Based on the above definition and the semantics of conjunctive grammars (see [Okh01]), we can now formally define the locally stratified semantics of boolean grammars:

Definition 10. *Let $G = (\Sigma, N, P, S)$ be a boolean grammar stratified by a canonical stratum-function f. Let n_i be the (unique) length of strings in the i-th stratum and N_i be the set of nonterminal symbols of this same stratum. The locally stratified semantics of G is the interpretation $L_G = \bigcup_{i=0}^{\infty} I_i$ where $I_0 = \perp$ and $I_{i+1}(A) = I_i(A) \cup \Delta_i(A)$, for every $A \in N$, and Δ_i is the interpretation that corresponds to the semantics of the conjunctive grammar $G_i = G/(I_i, N_i, n_i)$.*

We now demonstrate that the above construction is independent of the selection of the canonical stratum-function f.

In the following lemmas and theorem we denote by M_G the well-founded model of G (see [KNR06]) and by L_G the locally stratified semantics of G (see Definition 10). Furthermore, we treat grammars as function from $\Sigma^* \to \{0, 1\}$, ie., we write $I(C)(w) = 1$ rather than $w \in I(C)$. The proofs of the following lemmas are rather lengthy and will appear in the full version of the paper.

Lemma 1. *Let $G = (\Sigma, N, P, S)$ be a locally stratified boolean grammar. Then, for every $C \in N$ and for every $w \in \Sigma^*$ if $L_G(C)(w) = 1$ then $M_G(C)(w) = 1$.*

Lemma 2. *Let $G = (\Sigma, N, P, S)$ be a locally stratified boolean grammar. Then, for every $C \in N$ and for every $w \in \Sigma^*$, if $M_G(C)(w) = 1$ then $L_G(C)(w) = 1$.*

Lemma 3. *Let $G = (\Sigma, N, P, S)$ be a locally stratified boolean grammar. Then, for every $C \in N$ and for every $w \in \Sigma^*$, $M_G(C)(w) \in \{0, 1\}$ (ie., the well-founded model of a locally stratified boolean grammar is total).*

Theorem 2. *Let $G = (\Sigma, N, P, S)$ be a boolean grammar that is locally stratified. Then, $L_G = M_G$ and therefore L_G is independent of the choice of the canonical stratum function.*

Proof. From Lemmas 1, 2, 3, we conclude that for every $C \in N$ and for every $w \in \Sigma^*$, $L_G(C)(w) = M_G(C)(w)$, i.e. the locally stratified semantics is independent of the choice of the stratum function and coincides with the well founded semantics, which happens to be two-valued in the case of locally stratified grammars. □

Notice that, since $L_G = M_G$, by the results in [KNR06] we conclude that L_G is the least fixed point of an appropriate operator Ω_G assosiated with the grammar G.

5 An Application of the Locally Stratified Semantics

In this section we demonstrate that the language $\{a^{2^n} \mid n \geq 0\}$ is expressible under the locally stratified semantics. Notice that it is not known whether this language can be captured by the stratified semantics [Wro05].

Proposition 1. *The language $\{a^{2^n} \mid n \geq 0\}$ is expressible under the locally stratified semantics.*

Proof. A boolean grammar for the language $\{a^{2^n} \mid n \geq 0\}$ is given in [Okh04] (it is the grammar we have adopted in Example 1). We show below how one can prove that this grammar defines the desired language under the locally-stratified semantics. We restrict attention to a part of the grammar that is needed in order to define the non-terminal X. The proof for the remaining parts of the grammar is actually similar or simpler and is omitted.

Consider the boolean grammar $G = (\Sigma, N, P, X)$, where $\Sigma = \{a\}$ and P contains the following rules:

$$X \to aCC \quad A \to \neg X \quad B \to \neg AA \quad C \to \neg BB$$

We show that G defines the language $L = \{a^n \mid \exists k \geq 0 : 2^{3k} \leq n \leq 2^{3k+2} - 1\}$. Grammar G is locally stratified by the canonical stratum-function f with $f(X, w) = 4 \cdot |w|$, $f(A, w) = 4 \cdot |w| + 1$, $f(B, w) = 4 \cdot |w| + 2$ and $f(C, w) = 4 \cdot |w| + 3$. We show that $L_G(X) = L$, $L_G(A) = L_1$, $L_G(B) = L_2$, $L_G(C) = L_3$, where:

$$L_1 = \{a^n \mid \exists k \geq 0 : 2^{3k+2} \leq n \leq 2^{3k+3} - 1\} \cup \{\epsilon\}$$
$$L_2 = \{a^n \mid \exists k \geq 0 : 2^{3k+1} - 1 \leq n \leq 2^{3k+2} - 1\}$$
$$L_3 = \{a^n \mid \exists k \geq 0 : 2^{3k} - 1 \leq n \leq 2^{3k+1} - 1\}$$

It is easy to check (see also [Lei94]) that $L = \{a\} \circ L_3 \circ L_3$, $L_1 = \Sigma^* - L$, $L_2 = \Sigma^* - (L_1 \circ L_1)$ and $L_3 = \Sigma^* - (L_2 \circ L_2)$. In order to prove our claim we will prove by induction that: $I_i(X) = L \cap \Sigma^{\leq \lfloor \frac{i-1}{4} \rfloor}$, $I_i(A) = L_1 \cap \Sigma^{\leq \lfloor \frac{i-2}{4} \rfloor}$, $I_i(B) = L_2 \cap \Sigma^{\leq \lfloor \frac{i-3}{4} \rfloor}$, and $I_i(C) = L_3 \cap \Sigma^{\leq \lfloor \frac{i-4}{4} \rfloor}$. The case $i = 0$ is trivial. Suppose that our claim holds for i. We will show that it also holds for $i+1$. Let $n = \lfloor \frac{i}{4} \rfloor$, i.e. n is the length of the string in the unique pair in stratum i. The proof is by a case analysis:

Case 1: $i \bmod 4 = 0$. By the induction hypothesis $I_i(X) = L \cap \Sigma^{\leq n-1}$, $I_i(A) = L_1 \cap \Sigma^{\leq n-1}$, $I_i(B) = L_2 \cap \Sigma^{\leq n-1}$, $I_i(C) = L_3 \cap \Sigma^{\leq n-1}$. We will show that: $I_{i+1}(X) = L \cap \Sigma^{\leq n}$, $I_{i+1}(A) = L_1 \cap \Sigma^{\leq n-1}$, $I_{i+1}(B) = L_2 \cap \Sigma^{\leq n-1}$, $I_{i+1}(C) = L_3 \cap \Sigma^{\leq n-1}$. Stratum i contains a unique pair (X, a^n), $N_i = \{X\}$ and P_i contains the rule $X \to D_1$. Suppose that $a^n \in L$. Then, $a^n \in (\{a\} \circ L_3 \circ L_3)$, i.e there exist m, r, with $0 \leq m, r \leq n-1$, such that $n = m + r - 1$ and $a^m, a^r \in L_3$. By the induction hypothesis $a^m, a^r \in I_i(C)$. Therefore, $a^n \in \hat{I}_i(aCC)$, which implies that the rule $D_1 \to a^n$ is in P_i. Consequently, $\Delta_i(X) = \{a^n\}$ and $I_{i+1}(X) = I_i \cup \{a^n\} = L \cap \Sigma^{\leq n}$. Suppose that $a^n \notin L$. Then, $a^n \notin (\{a\} \circ L_3 \circ L_3)$, which implies, using the induction hypothesis, that $a^n \notin \hat{I}_i(aCC)$. Thus, the rule $D_1 \to a^n$ is not in P_i. Consequently, $\Delta_i(X) = \emptyset$ and $I_{i+1}(X) = I_i \cup \emptyset = L \cap \Sigma^{\leq n}$. Moreover, $I_{i+1}(V) = I_i(V)$, for every $V \in \{A, B, C\}$. The inductive step is proved for Case 1.

Case 2: $i \bmod 4 = 1$. By the induction hypothesis $I_i(X) = L \cap \Sigma^{\leq n}$, $I_i(A) = L_1 \cap \Sigma^{\leq n-1}$, $I_i(B) = L_2 \cap \Sigma^{\leq n-1}$, $I_i(C) = L_3 \cap \Sigma^{\leq n-1}$. We will show that: $I_{i+1}(X) = L \cap \Sigma^{\leq n}$, $I_{i+1}(A) = L_1 \cap \Sigma^{\leq n}$, $I_{i+1}(B) = L_2 \cap \Sigma^{\leq n-1}$, $I_{i+1}(C) = L_3 \cap \Sigma^{\leq n-1}$. Stratum i contains a unique pair (A, a^n), $N_i = \{A\}$ and P_i contains the rule $A \to D_2$. Suppose that $a^n \in L_1$. Then, $a^n \notin L$. By the induction hypothesis $a^n \notin I_i(X)$. Therefore, $a^n \in \hat{I}_i(\neg X)$, which implies that the rule $D_2 \to a^n$ is in P_i. Consequently, $\Delta_i(A) = \{a^n\}$ and $I_{i+1}(A) = I_i \cup \{a^n\} = L_1 \cap \Sigma^{\leq n}$. Suppose that $a^n \notin L_1$. Then, $a^n \in L$. By the induction hypothesis $a^n \in I_i(X)$. Therefore, $a^n \notin \hat{I}_i(\neg X)$, which implies that the rule $D_2 \to a^n$ is not in P_i. Consequently, $\Delta_i(A) = \emptyset$ and $I_{i+1}(A) = I_i \cup \emptyset = L_1 \cap \Sigma^{\leq n}$. Moreover, $I_{i+1}(V) = I_i(V)$, for every $V \in \{X, B, C\}$. The inductive step is proved for Case 2.

Case 3: $i \bmod 4 = 2$. By the induction hypothesis $I_i(X) = L \cap \Sigma^{\leq n}$, $I_i(A) = L_1 \cap \Sigma^{\leq n}$, $I_i(B) = L_2 \cap \Sigma^{\leq n-1}$, $I_i(C) = L_3 \cap \Sigma^{\leq n-1}$. We show that: $I_{i+1}(X) = L \cap \Sigma^{\leq n}$, $I_{i+1}(A) = L_1 \cap \Sigma^{\leq n}$, $I_{i+1}(B) = L_2 \cap \Sigma^{\leq n}$, $I_{i+1}(C) = L_3 \cap \Sigma^{\leq n-1}$. Stratum i contains a unique pair (B, a^n), $N_i = \{B\}$ and P_i contains the rule $B \to D_3$. Suppose that $a^n \in L_2$. Then, $a^n \notin (L_1 \circ L_1)$, which by the induction hypothesis implies that $a^n \in \hat{I}_i(\neg AA)$. Therefore, the rule $D_3 \to a^n$ is in P_i. Consequently, $\Delta_i(B) = \{a^n\}$ and $I_{i+1}(B) = I_i(B) \cup \{a^n\} = L_2 \cap \Sigma^{\leq n}$. Suppose that $a^n \notin L_2$. Then, $a^n \in (L_1 \circ L_1)$, ie., there exist m, r, $0 \leq m, r \leq n$, such that $n = m + r$ and $a^m, a^r \in L_1$. By the induction hypothesis $a^m, a^r \in I_i(A)$. Therefore, $a^n \notin \hat{I}_i(\neg AA)$, which implies that the rule $D_3 \to a^n$ is not in P_i. Therefore, $\Delta_i(B) = \emptyset$ and $I_{i+1}(B) = I_i(B) \cup \emptyset = L_2 \cap \Sigma^{\leq n}$. Moreover, $I_{i+1}(V) = I_i(V)$, for every $V \in \{X, A, C\}$. The inductive step is proved for Case 3.

Case 4: $i \bmod 4 = 3$. This case is similar to Case 3.

6 Beyond Local Stratification

We have defined the class of locally stratified boolean grammars and have demonstrated that they have a well-defined semantics. The class of locally stratified grammars is broader than that of the stratified ones (see Corollary 2) and can express a language which is not known to be captured by the stratified semantics. Additionally, the locally stratified boolean grammars can be considered as "well-behaved" and useful for applications, since their well-founded semantics is total.

There exist however certain natural and useful boolean grammars that fail to be locally stratified. Consider the following modified version of our motivating example in Section 3:

$$\begin{array}{ll} Even & \to \epsilon \\ Even & \to A\ Odd \\ Odd & \to \neg Even \\ A & \to a \end{array}$$

Despite its obvious equivalence with the initial grammar, the above grammar is not locally stratified since its skeleton in not stratified. This same phenomenon occurs in a slightly different form in the area of logic programming. For example, consider the following (locally stratified) logic program:

$$\begin{array}{ll} p & \leftarrow q(b). \\ q(a) & \leftarrow \neg p. \end{array}$$

This program can be written in an equivalent way as:

$$\begin{array}{lll} p & \leftarrow & equal(X, b), q(X). \\ q(a) & \leftarrow & \neg p. \\ equal(X, X). & & \end{array}$$

Despite the fact that the two programs above are equivalent from a semantic point of view, the second program *is not* locally stratified. The reasons that lead to the above problem are not hard to detect. Local stratification is a purely syntactical notion, while the above phenomenon requires a more in-depth (namely semantical) inspection of the grammar (or logic program).

Therefore, given a boolean grammar, if we would like to use some more powerful notion than local stratification, we would have to perform some kind of semantic analysis regarding the grammar under consideration. For example, consider again the grammar given in the beginning of this section. The decision procedure we introduced in Section 3 fails for this program because we don't know whether the nonterminal A can produce the empty string or not. In this particular example this can be checked by an easy inspection of the rules of the grammar. However, in other more complicated cases this would require a more in-depth inspection of the rules of the grammar. It is possible that based on semantic information one could define more general and effective tests, but it is not immediately obvious how far this approach can take us.

Closing, we believe that another very important subject for future research is the classification of the expressive power of the various semantics for boolean grammars (namely the work presented in this paper compared to the approaches of [Okh04,Wro05,KNR06]). We believe it is possible that certain approaches define separate classes of languages, but such an investigation seems to require new tools to be employed.

References

[ABW88] Apt, K., Bol, R., Walker, A.: Towards a Theory of Declarative Knowledge. In J. Minker, editor, Foundations of Deductive Databases and Logic Programming, Ed. Morgan Kaufmann, Los Altos, CA, 89–148.

[CB94] Cholak, P., Blair, H.: The Complexity of Local Stratification. Fundamenta Informaticae **21(4)** (1994) 333–344.

[KNR06] Kountouriotis, V., Nomikos, C., Rondogiannis, P.: Well-Founded Semantics of Boolean Grammars. DLT (2006) 203–214.

[Lei94] Leiss, E. L.: Unrestricted Complementation in Language Equations over a One-Letter Alphabet. Theoretical Computer Science. **132 (2)** (1994) 71–84.

[NRG05] Nomikos, C., Rondogiannis, P., Gergatsoulis, M.: Temporal Stratification Tests for Linear and Branching-Time Deductive Databases. Theoretical Computer Science. **342(2-3)** (2005) 382–415.

[Okh01] Okhotin, A.: Conjunctive Grammars. Journal of Automata, Languages and Combinatorics **6(4)** (2001) 519–535.

[Okh04] Okhotin, A.: Boolean Grammars. Information and Computation **194(1)** (2004) 19–48.

[PP90] Przymusinska, H., Przymusinski, T.: Semantic Issues in Deductive Databases and Logic Programs. In R. Banerji, editor, Formal Techniques in Artificial Intelligence: a Source-Book, North Holland (1990) 321–367.

[Prz88] Przymusinski, T., C,: On the Declarative Semantics of Deductive Databases and Logic Programs. In J. Minker, editor, Foundations of Deductive Databases and Logic Programming, Ed. Morgan Kaufmann, Los Altos, CA, 193–216.

[Ron01] Rondogiannis, P.: Stratified Negation in Temporal Logic Programming and the Cycle-Sum Test. Theoretical Computer Science. **254 (1-2)** (2001) 663–676.

[RW05] Rondogiannis, P., Wadge, W., W.: Minimum Model Semantics for Logic Programs with Negation-as-Failure. ACM Transactions on Computational Logic **6(2)** (2005) 441–467.

[vGRS91] van Gelder, A, Ross, K., A., Schlipf, J., S.: The Well-Founded Semantics for General Logic Programs. Journal of the ACM **38(3)** (1991) 620–650.

[Wro05] Wrona M.: Stratified Boolean Grammars. MFCS (2005) 801–812.