ELSEVIER Theoretical Computer Science 290 (2003) 531–544

Theoretical Computer Science

www.elsevier.com/locate/tcs

Parsing as abstract interpretation of grammar semantics

Patrick Cousot[a,*], Radhia Cousot[b]

[a] *Département d'informatique, École Normale Supérieure, 45 rue d'Ulm, 75230 Paris Cedex 05, France*
[b] *Laboratoire d'informatique, CNRS & École Polytechnique, 91128 Palaiseau Cedex, France*

Received 15 January 2001; accepted 16 August 2001
Communicated by G. Levi

Abstract

Earley's parsing algorithm is shown to be an abstract interpretation of a refinement of the derivation semantics of context-free grammars. © 2002 Elsevier Science B.V. All rights reserved.

1. Introduction

Abstract interpretation is a theory of the approximation of the mathematical structures involved in the formalization of the semantics of computer systems [6]. It offers a unifying point of view on *static program analysis* [4] (including *data flow analysis* [6,8] and *typing* [2]) of specification and programming languages, *model-checking* [8], etc. Following this synthetic point of view, we show that Earley's parsing algorithm [9] can be formally designed by abstract interpretation of a refinement of the derivation semantics of context-free grammars.

2. Context-free grammars, derivations, generated language and parsing

The set of *finite words* on an alphabet $\mathscr{A}$ is denoted $\mathscr{A}^\star$. This includes the empty word ε. A *language* on the alphabet $\mathscr{A}$ is a subset of $\mathscr{A}^\star$. A *context-free grammar* $\mathscr{G}$ is a quadruple $\langle \mathscr{N}, \mathscr{T}, \mathscr{P}, A \rangle$, where

- $X, Y, \ldots \in \mathscr{N}$ is the finite set of *nonterminals*;

---

[*] Corresponding author.
E-mail addresses: cousot@ens.fr (P. Cousot), rcousot@lix.polytechnique.fr (R. Cousot).

PII: S0304-3975(02)00034-8

- the distinguished nonterminal $A \in \mathscr{N}$ is the *axiom*;
- $a, b, \ldots \in \mathscr{T}$, such that $\mathscr{T} \cap \mathscr{N} = \emptyset$, is the finite set of *terminals*;
- $\mathscr{V} \triangleq (\mathscr{N} \cup \mathscr{T}) \setminus \{A\}$ is the *vocabulary*;
- $\alpha, \beta, \ldots \in \mathscr{V}^{\star}$ is the set of finite words on the vocabulary $\mathscr{V}$;
- $\mathscr{P} \subseteq \mathscr{N} \times \mathscr{V}^{\star}$ is the finite set of *productions*, $\langle X, \alpha \rangle \in \mathscr{P}$ being written $X \overset{\mathscr{G}}{\rightarrow} \alpha$.

Observe that the axiom A cannot appear on the right-hand side α of productions $\langle X, \alpha \rangle$. This restriction can be easily bypassed by introducing a new axiom A' such that $A' \overset{\mathscr{G}}{\rightarrow} A$.

The *semantics* of a grammar $\mathscr{G}$ can be defined as the *derivation relation* $\overset{\mathscr{G}}{\Rightarrow}$ which is the least relation such that a nonterminal derives to the right-hand side of any of its productions, as specified by the following axiom schema ($X \in \mathscr{N}$, $\alpha \in \mathscr{V}^{\star}$):

$$X \overset{\mathscr{G}}{\Rightarrow} \alpha \quad \text{whenever } X \overset{\mathscr{G}}{\rightarrow} \alpha \tag{1}$$

and a word derives to another word by replacement of a nonterminal by any one of its derivations, as specified by the following inference rule schema:

$$\frac{X \overset{\mathscr{G}}{\Rightarrow} \alpha Y \gamma, \quad Y \overset{\mathscr{G}}{\Rightarrow} \beta}{X \overset{\mathscr{G}}{\Rightarrow} \alpha \beta \gamma}, \quad X, Y \in \mathscr{N}, \ \alpha, \beta, \gamma \in \mathscr{V}^{\star}. \tag{2}$$

The *leftmost derivation* $\overset{\mathscr{G}}{\Rightarrow}_{\ell}$ is defined in the same way but for the nonterminal replacement which is restricted to the leftmost nonterminal:

$$X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \alpha, \quad \text{whenever } X \overset{\mathscr{G}}{\rightarrow} \alpha \tag{3}$$

$$\frac{X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \alpha Y \gamma, \quad Y \overset{\mathscr{G}}{\Rightarrow}_{\ell} \beta}{X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \alpha \beta \gamma}, \quad X, Y \in \mathscr{N}, \ \alpha \in \mathscr{T}^{\star}, \ \beta, \gamma \in \mathscr{V}^{\star} \tag{4}$$

Similarly, the *leftmost derivation from the axiom* $\overset{\mathscr{G}}{\Rightarrow}_{A,\ell}$ is the restriction of the leftmost derivation $\overset{\mathscr{G}}{\Rightarrow}_{\ell}$ to nonterminals deriving from the grammar axiom:

$$A \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha, \quad \text{whenever } A \overset{\mathscr{G}}{\rightarrow} \alpha \tag{5}$$

$$\frac{X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha Y \gamma}{Y \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \beta}, \quad X, Y \in \mathscr{N}, \ \alpha \in \mathscr{T}^{\star}, \ \gamma \in \mathscr{V}^{\star}, \ Y \overset{\mathscr{G}}{\rightarrow} \beta \tag{6}$$

$$\frac{X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha Y \gamma, \quad Y \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \beta}{X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha \beta \gamma}, \quad X, Y \in \mathscr{N}, \ \alpha \in \mathscr{T}^{\star}, \ \beta, \gamma \in \mathscr{V}^{\star} \tag{7}$$

The *language* $\mathscr{L}_{\mathscr{G}}$ *generated by a grammar* $\mathscr{G}$ is the set of terminal words deriving from the axiom A:

$$\mathscr{L}_{\mathscr{G}} \triangleq \{\alpha \in \mathscr{T}^{\star} \mid A \overset{\mathscr{G}}{\Rightarrow} \alpha\}. \tag{8}$$

Equivalently, the language generated by a grammar can be defined using the leftmost derivation [12, Theorem 4.1.1]:

$$\mathscr{L}_{\mathscr{G}} = \{\alpha \in \mathscr{T}^{\star} \mid A \overset{\mathscr{G}}{\Rightarrow}_{\ell} \alpha\}. \tag{9}$$

Equivalently, we can also use the leftmost derivation from the axiom:

Lemma 1.

$$\mathscr{L}_{\mathscr{G}} = \{\alpha \in \mathscr{T}^{\star} \mid A \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha\}. \tag{10}$$

Proof. Obviously, if we have proved $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha$ we can prove $X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \alpha$ using (3) for either (5) or (6) and (4) for (7).

Reciprocally, we prove that if $X = A$ or $\exists \eta \in \mathscr{T}^{\star}$, $\xi \in \mathscr{V}^{\star}$: $A \overset{\mathscr{G}}{\Rightarrow}_{\ell} \eta X \xi$ and $X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \delta$ then $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \delta$.

The proof is on the length of the proof of $X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \delta$ by the formal system (3)–(4).

- if we have proved $X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \delta$ by (5) then $X \overset{\mathscr{G}}{\rightarrow} \delta$ and there are two subcases
  - if $X = A$ then $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \delta$ follows from (5);
  - otherwise, there exist $\eta \in \mathscr{T}^{\star}$ and $\xi \in \mathscr{V}^{\star}$ such that $A \overset{\mathscr{G}}{\Rightarrow}_{\ell} \eta X \xi$. So, by induction, we can prove that $A \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \eta X \xi$ whence $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \delta$ by (6).
- otherwise, we have proved $X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \delta$ by (4) so we have $\delta = \alpha\beta\gamma$, $\alpha \in \mathscr{T}^{\star}$ and we made subproofs for $X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \alpha Y \gamma$ and $Y \overset{\mathscr{G}}{\Rightarrow}_{\ell} \beta$. There are now two subcases:
  - if $X = A$ then by induction $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha Y \gamma$ that is $A \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha Y \gamma$ with $\alpha \in \mathscr{T}^{\star}$ so that again by induction $Y \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \beta$. By (7), we conclude that $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha\beta\gamma$ that is $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \delta$;
  - otherwise, there exist $\eta \in \mathscr{T}^{\star}$ and $\xi \in \mathscr{V}^{\star}$ such that $A \overset{\mathscr{G}}{\Rightarrow}_{\ell} \eta X \xi$. By $X \overset{\mathscr{G}}{\Rightarrow}_{\ell} \alpha Y \gamma$ and (4), it follows that $A \overset{\mathscr{G}}{\Rightarrow}_{\ell} \eta \alpha Y \gamma \xi$ with $\eta\alpha \in \mathscr{T}^{\star}$. Hence we can apply the induction hypothesis and therefore prove that $Y \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \beta$. By (7), we conclude that $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha\beta\gamma$, whence $X \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \delta$.

We conclude that $A \overset{\mathscr{G}}{\Rightarrow}_{\ell} \alpha$ if and only if $A \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \alpha$ so that (9) implies (10). □

Parsing of a given terminal word $\omega \in \mathscr{T}^{\star}$ for a given grammar $\mathscr{G}$ consists in deciding whether this word ω belongs to the language generated by the grammar $\mathscr{G}$: $\omega \in \mathscr{L}_{\mathscr{G}}$.

3. Fixpoint semantics of formal systems

It is well-known that formal systems specify a least fixpoint [1,7]. The axioms and rule schemata of a formal system are interpreted as rule instances $\Phi \triangleq \{P_i/c_i \mid i \in \Delta\}$ on a given *universe* $\mathscr{U}$ where for all $i \in \Delta$, $P \subseteq \mathscr{U}$ is the *premise* (which is the empty set $\emptyset$ for axiom instances) and $c_i \in \mathscr{U}$ is the *conclusion* of the *rule instance* P_i/c_i. The subset of the universe $\mathscr{U}$ specified by the formal system Φ is defined as its semantics

$[\![\Phi]\!] \triangleq \mathrm{lfp}^{\subseteq} F_\Phi$ where the *consequence operator*

$$F_\Phi(X) \triangleq \{c_i \mid i \in \Delta \wedge P_i \subseteq X\} \tag{11}$$

is the set of valid consequences of the hypothesis X. The consequence operator F_Φ on $\wp(\mathscr{U})$ is $\subseteq$-monotonic so that the least fixpoint $\mathrm{lfp}^{\subseteq} F_\Phi$ does exist [13]. The fixpoint semantics is equivalent to the more traditional one based on *formal proofs* [1].

For example, the inference system (5)–(7) defines the leftmost derivation from the grammar axiom as

$$\overset{\mathscr{G}}{\Rightarrow}_{A,\ell} = \mathrm{lfp}^{\subseteq} \mathscr{D}^{\mathscr{G}}_{A,\ell},$$

$$\begin{aligned}\mathscr{D}^{\mathscr{G}}_{A,\ell}(R) \triangleq {} & \{\langle A, \alpha\rangle \mid A \overset{\mathscr{G}}{\rightarrow} \alpha\} \\ & \cup \{\langle Y, \beta\rangle \mid \langle X, \alpha Y \gamma\rangle \in R \wedge \alpha \in \mathscr{T}^\star \wedge Y \overset{\mathscr{G}}{\rightarrow} \beta\} \\ & \cup \{\langle X, \alpha\beta\gamma\rangle \mid \langle X, \alpha Y\gamma\rangle \in R \wedge \alpha \in \mathscr{T}^\star \wedge \langle Y, \beta\rangle \in R\}.\end{aligned} \tag{12}$$

4. Earley's parsing algorithm

4.1. Earley's items

Given a terminal word $\omega \in \mathscr{T}^\star$, $\omega = \omega_1 \ldots \omega_n$, $n \geqslant 0$ (which is ε when $n = 0$), *Earley's parsing algorithm* [9,11] involves *Earley's items* which are quintuples written as follows:

$$\langle X \rightarrow \alpha \cdot \beta, i, j\rangle,$$

where $X \overset{\mathscr{G}}{\rightarrow} \alpha\beta$ is a production of the given grammar $\mathscr{G}$ and $0 \leqslant i \leqslant j \leqslant n$. A *valid Earley's item* is an assertion or judgement stating that $\alpha \overset{\mathscr{G}}{\Rightarrow} \omega_{i+1} \ldots \omega_j$ (that is $\alpha \overset{\mathscr{G}}{\Rightarrow} \varepsilon$ when $i = j$). Valid Earley's items are derived left to right and top-down starting from the grammar axiom. The set $\mathscr{I}^{\mathrm{E}}_{\mathscr{G},\omega}$ of valid Earley's items for the grammar $\mathscr{G}$ and input word ω is specified by the formal system (13)–(16) below.

4.2. Rule-based specification of Earley's parsing algorithm

The *initialization axioms* are instances of the following schema (for all productions $A \overset{\mathscr{G}}{\rightarrow} \gamma$ of the grammar axiom A):

$$\langle A \rightarrow \cdot\gamma, 0, 0\rangle. \tag{13}$$

The *derivation rules* are instances of the following schema (for all productions $X \overset{\mathscr{G}}{\rightarrow} \alpha Y \beta$ and $Y \overset{\mathscr{G}}{\rightarrow} \gamma$ of the grammar $\mathscr{G}$ and $0 \leqslant i \leqslant j \leqslant n$):

$$\frac{\langle X \rightarrow \alpha \cdot Y\beta, i, j\rangle}{\langle Y \rightarrow \cdot\gamma, j, j\rangle}. \tag{14}$$

The *reduction rule schema* is (for all productions $X \overset{\mathscr{G}}{\longrightarrow} \alpha Y\beta$ and $Y \overset{\mathscr{G}}{\longrightarrow} \gamma$ of the grammar $\mathscr{G}$ and $0 \leqslant k \leqslant i \leqslant j \leqslant n$)

$$\frac{\langle X \to \alpha \cdot Y\beta, k, i\rangle, \quad \langle Y \to \gamma\cdot, i, j\rangle}{\langle X \to \alpha Y \cdot \beta, k, j\rangle}. \tag{15}$$

The *advance rule schema* is (for all productions $X \overset{\mathscr{G}}{\longrightarrow} \alpha a\beta$ of the grammar $\mathscr{G}$ and $0 \leqslant i < j \leqslant n$ such that $a = \omega_j$)

$$\frac{\langle X \to \alpha \cdot \omega_j\beta, i, j-1\rangle}{\langle X \to \alpha\omega_j \cdot \beta, i, j\rangle}. \tag{16}$$

The parsing succeeds, that is $\omega \in \mathscr{L}_{\mathscr{G}}$, if and only if one can derive a final Earley's item of the form $\langle A \to \gamma\cdot, 0, n\rangle$ where A is the grammar axiom.

4.3. Fixpoint specification of Earley's parsing algorithm

The derivation of the set $\mathscr{I}^{\mathrm{E}}_{\mathscr{G},\omega}$ of valid Earley's items by the formal system (13)–(16) consists in computing the least fixpoint

$$\mathscr{I}^{\mathrm{E}}_{\mathscr{G},\omega} \triangleq \mathrm{lfp}^{\subseteq}\, \mathscr{F}^{\mathrm{E}}_{\mathscr{G},\omega},$$

$$\begin{aligned}
\mathscr{F}^{\mathrm{E}}_{\mathscr{G},\omega}(I) \triangleq\ & \{\langle A \to \cdot\gamma, 0, 0\rangle \mid A \overset{\mathscr{G}}{\longrightarrow} \gamma\} \\
& \cup \{\langle Y \to \cdot\gamma, j, j\rangle \mid \langle X \to \alpha \cdot Y\beta, i, j\rangle \in I\} \\
& \cup \{\langle X \to \alpha Y \cdot \beta, k, j\rangle \mid \langle X \to \alpha \cdot Y\beta, k, i\rangle \in I \wedge \langle Y \to \gamma\cdot, i, j\rangle \in I\} \\
& \cup \{\langle X \to \alpha\omega_j \cdot \beta, i, j\rangle \mid \langle X \to \alpha \cdot \omega_j\beta, i, j-1\rangle \in I\}.
\end{aligned} \tag{17}$$

The Earley's parsing algorithm [9] terminates by checking that a final item is valid, so that the correctness of the original algorithm and its variants can be specified as

$$\omega \in \mathscr{L}_{\mathscr{G}} \quad \Leftrightarrow \quad \langle A \to \gamma\cdot, 0, n\rangle \in \mathscr{I}^{\mathrm{E}}_{\mathscr{G},\omega}. \tag{18}$$

5. Elements of abstract interpretation

5.1. The abstraction

The approximation or abstraction of a semantics is specified by a *Galois connection* [6] that is a pair of maps $\boldsymbol{\alpha} \in L \mapsto M$ and $\boldsymbol{\gamma} \in M \mapsto L$ between posets $\langle L, \leqslant\rangle$ and $\langle M, \sqsubseteq\rangle$ satisfying $\forall x \in L$: $\forall y \in M$: $\boldsymbol{\alpha}(x) \sqsubseteq y \Leftrightarrow x \leqslant \boldsymbol{\gamma}(y)$ which is written $\langle L, \leqslant\rangle \underset{\boldsymbol{\alpha}}{\overset{\boldsymbol{\gamma}}{\leftrightarrows}} \langle M, \sqsubseteq\rangle$.

An equivalent definition is $\boldsymbol{\alpha} \in L \mapsto M$ and $\boldsymbol{\gamma} \in M \mapsto L$ are monotonic, $\boldsymbol{\alpha} \circ \boldsymbol{\gamma} \mathrel{\dot{\sqsubseteq}} \mathbf{1}_M$ and $\mathbf{1}_L \mathrel{\dot{\leqslant}} \boldsymbol{\gamma} \circ \boldsymbol{\alpha}$ where $f \mathrel{\dot{\preccurlyeq}} g$ is the pointwise extension of $\preccurlyeq$ that is $\forall x \in L$: $f(x) \preccurlyeq g(x)$ and $\mathbf{1}_S$ is the identity map $\forall x \in S$: $\mathbf{1}_S(x) = x$ on the set S.

We will use the fact that if $\boldsymbol{\alpha}$ preserves least upper bounds existing in $\langle L, \leqslant\rangle$ then it has a unique adjoint $\boldsymbol{\gamma}$ such that $\langle L, \leqslant\rangle \underset{\boldsymbol{\alpha}}{\overset{\boldsymbol{\gamma}}{\leftrightarrows}} \langle M, \sqsubseteq\rangle$.

5.2. *The abstract interpretation of the semantics*

If $\langle L, \leqslant\rangle$ is a complete lattice and $f \in L \mapsto L$ is a monotone map on L, then it has a least fixpoint $\mathrm{lfp}^{\leqslant} f$ [13] which is interpreted as a *concrete semantics*. The monotone map $g \in M \mapsto M$ on M is a said to be a *locally complete abstraction* of f if and only if $\boldsymbol{\alpha} \circ f = g \circ \boldsymbol{\alpha}$ (see [6, 7.1.0.4(3)]). This implies *fixpoint completeness* in that the *abstract semantics* $\mathrm{lfp}^{\leqslant} g = \boldsymbol{\alpha}(\mathrm{lfp}^{\leqslant} f)$ is the precise or exact abstraction of the concrete semantics $\mathrm{lfp}^{\leqslant} f$ by the abstraction function $\boldsymbol{\alpha}$.

Lemma 2. *If $\langle L, \leqslant\rangle$ is a complete lattice, $\langle L, \leqslant\rangle \underset{\boldsymbol{\alpha}}{\overset{\boldsymbol{\gamma}}{\leftrightarrows}} \langle M, \sqsubseteq\rangle$, $f \in L \mapsto L$ and $g \in M \mapsto M$ are monotone maps and $\boldsymbol{\alpha} \circ f = g \circ \boldsymbol{\alpha}$ then $\boldsymbol{\alpha}(\mathrm{lfp}^{\leqslant} f) = \mathrm{lfp}^{\sqsubseteq} g$.*

Proof. $\boldsymbol{\alpha} \circ f \circ \boldsymbol{\gamma} = g \circ \boldsymbol{\alpha} \circ \boldsymbol{\gamma} \,\dot{\sqsubseteq}\, g$ by monotony and $\boldsymbol{\alpha}(\mathrm{lfp}^{\leqslant} f) = \mathrm{lfp}^{\sqsubseteq} g$ by [6, 7.1.0.4(3)]. □

Numerous examples of locally complete abstractions of the derivation semantics of context-free grammars are given in [3]. In this paper, we show that parsing is another one.

6. Concrete grammar item semantics

Our task is now to show that Earley's parsing algorithm (17) is an abstract interpretation of the grammar semantics. We consider a refinement of the leftmost derivation from the axiom semantics (12) in order to take into account the possible contexts of derivations.

6.1. *Grammar items*

The grammar semantics defines *grammar items* which are quintuples written

$$[\lambda, X \longrightarrow \alpha \cdot \beta, \gamma],$$

where $\lambda, \gamma \in \mathscr{T}^{\star}$ and $X \overset{\mathscr{G}}{\longrightarrow} \alpha\beta$. The interpretation of a valid grammar item is that there exists $\eta \in \mathscr{V}^{\star}$ such that $A \overset{\mathscr{G}}{\Rightarrow} \lambda X \eta$, $X \overset{\mathscr{G}}{\longrightarrow} \alpha\beta$ and $\alpha \overset{\mathscr{G}}{\Rightarrow} \gamma$.

The set $\mathscr{I}_{\mathscr{G}}$ of *valid grammar items* is defined by the formal system (19)–(22) below.

6.2. *Rule-based specification of the grammar item semantics*

The *initialization axiom schema* is (for all productions $A \overset{\mathscr{G}}{\longrightarrow} \beta$ of the grammar axiom A)

$$[\varepsilon, A \longrightarrow \cdot \beta, \varepsilon]. \tag{19}$$

The *derivation rule schema* is (for all productions $X \overset{\mathscr{G}}{\to} \alpha Y\beta$ and $Y \overset{\mathscr{G}}{\to} \delta$ of the grammar $\mathscr{G}$)

$$\frac{[\lambda, X \to \alpha \cdot Y\beta, \gamma]}{[\lambda\gamma, Y \to \cdot\delta, \varepsilon]}. \tag{20}$$

The *reduction rule schema* is (for all productions $X \overset{\mathscr{G}}{\to} \alpha Y\beta$ and $Y \overset{\mathscr{G}}{\to} \gamma$ of the grammar $\mathscr{G}$)

$$\frac{[\lambda, X \to \alpha \cdot Y\beta, \gamma], \quad [\lambda\gamma, Y \to \delta\cdot, \xi]}{[\lambda, X \to \alpha Y \cdot \beta, \gamma\xi]}. \tag{21}$$

The *advance rule schema* is (for all productions $X \overset{\mathscr{G}}{\to} \alpha a\beta$ of the grammar $\mathscr{G}$)

$$\frac{[\lambda, X \to \alpha \cdot a\beta, \gamma]}{[\lambda, X \to \alpha a \cdot \beta, \gamma a]}. \tag{22}$$

The derivation context from the axiom is always empty since the axiom never appears in the right-hand side of production:

Lemma 3. *If* $[\lambda, A \to \alpha \cdot \beta, \gamma] \in \mathscr{I}_{\mathscr{G}}$ *then* $\lambda = \varepsilon$.

Proof. We proceed by induction on the length of the proof that $[\lambda, A \to \alpha \cdot \beta, \gamma] \in \mathscr{I}_{\mathscr{G}}$ using (19)–(22).

This is obvious for the basis by (19). For the induction step, we cannot conclude the proof with (20) because there would be a grammar production of the form $\langle X, \alpha A \alpha \rangle$. So the proof ends with the use of either (21) or (22) and in both cases $\lambda = \varepsilon$ follows by induction. □

6.3. Fixpoint specification of the grammar item semantics

In fixpoint form, the grammar item semantics is

$$\mathscr{I}_{\mathscr{G}} = \operatorname{lfp}^{\subseteq} \mathscr{F}_{\mathscr{G}},$$

$$\begin{aligned}
\mathscr{F}_{\mathscr{G}}(I) \triangleq {} & \{[\varepsilon, A \to \cdot\beta, \varepsilon] \mid A \overset{\mathscr{G}}{\to} \beta\} \\
& \cup \{[\lambda\gamma, Y \to \cdot\delta, \varepsilon] \mid [\lambda, X \to \alpha \cdot Y\beta, \gamma] \in I \wedge Y \overset{\mathscr{G}}{\to} \delta\} \\
& \cup \{[\lambda, X \to \alpha Y \cdot \beta, \gamma\xi] \mid [\lambda, X \to \alpha \cdot Y\beta, \gamma] \in I \wedge [\lambda\gamma, Y \to \delta\cdot, \xi] \in I\} \\
& \cup \{[\lambda, X \to \alpha a \cdot \beta, \gamma a] \mid [\lambda, X \to \alpha \cdot a\beta, \gamma] \in I\}.
\end{aligned} \tag{23}$$

7. The leftmost derivation from the axiom is a complete abstraction of the grammar item semantics

7.1. The abstraction

We consider the elementwise abstraction

$$\boldsymbol{\alpha}^{\ell}(I) \triangleq \{\langle X, \gamma\beta\rangle \mid \exists \lambda \in \mathscr{T}^{\star}\colon\ [\lambda, X \to \alpha \cdot \beta, \gamma] \in I\}. \tag{24}$$

$\boldsymbol{\alpha}^{\ell}$ is a complete $\cup$-morphism so it is the lower adjoint of a Galois connection

$$\langle \wp(\mathscr{T}^{\star} \times \mathscr{N} \times \mathscr{V}^{\star} \times \mathscr{V}^{\star} \times \mathscr{T}^{\star}), \subseteq\rangle \underset{\boldsymbol{\alpha}^{\ell}}{\overset{\boldsymbol{\gamma}^{\ell}}{\leftrightarrows}} \langle \wp(\mathscr{N} \times \mathscr{V}^{\star}), \subseteq\rangle. \tag{25}$$

7.2. The abstract interpretation of the semantics

The leftmost derivation from the axiom semantics is a complete abstract interpretation of the grammar item semantics:

Lemma 4.

$$\overset{\mathscr{G}}{\Rightarrow}_{A,\ell} = \boldsymbol{\alpha}^{\ell}(\mathscr{I}_{\mathscr{G}}). \tag{26}$$

Proof.

$$\begin{aligned}
&\boldsymbol{\alpha}^{\ell} \circ \mathscr{F}_{\mathscr{G}}(I)\\
&= \wr\text{by def. (24) of } \boldsymbol{\alpha}^{\ell} \text{ and (23) of } \mathscr{F}_{\mathscr{G}}\,\wr\\
&\quad \{\langle A, \beta\rangle \mid A \overset{\mathscr{G}}{\to} \beta\}\\
&\quad \cup \{\langle Y, \delta\rangle \mid \exists \lambda \in \mathscr{T}^{\star}, \alpha, \beta \in \mathscr{V}^{\star}\colon\ [\lambda, X \to \alpha \cdot Y\beta, \gamma] \in I \wedge Y \overset{\mathscr{G}}{\to} \delta\}\\
&\quad \cup \{\langle X, \gamma\xi\beta\rangle \mid \exists \lambda \in \mathscr{T}^{\star}\colon\ [\lambda, X \to \alpha \cdot Y\beta, \gamma] \in I \wedge [\lambda\gamma, Y \to \delta\cdot, \xi] \in I\}\\
&\quad \cup \{\langle X, \gamma a\beta\rangle \mid \exists \lambda \in \mathscr{T}^{\star}\colon\ [\lambda, X \to \alpha \cdot a\beta, \gamma] \in I\}.\\
&= \wr\text{by def. (24) of } \boldsymbol{\alpha}^{\ell} \text{ so that } \exists \lambda \in \mathscr{T}^{\star}\colon\ [\lambda, X \to \alpha \cdot \beta, \gamma] \in I \text{ if and only if}\\
&\quad \langle X, \gamma\beta\rangle \in \boldsymbol{\alpha}^{\ell}(I) \text{ and } \gamma \in \mathscr{T}^{\star}\,\wr\\
&\quad \{\langle A, \beta\rangle \mid A \overset{\mathscr{G}}{\to} \beta\}\\
&\quad \cup \{\langle Y, \delta\rangle \mid \langle X, \gamma Y\beta\rangle \in \boldsymbol{\alpha}^{\ell}(I) \wedge \gamma \in \mathscr{T}^{\star} \wedge Y \overset{\mathscr{G}}{\to} \delta\}\\
&\quad \cup \{\langle X, \gamma\xi\beta\rangle \mid \langle X, \gamma Y\beta\rangle \in \boldsymbol{\alpha}^{\ell}(I) \wedge \gamma \in \mathscr{T}^{\star} \wedge \langle Y, \xi\rangle \in \boldsymbol{\alpha}^{\ell}(I)\}\\
&\quad \cup \{\langle X, \gamma a\beta\rangle \mid \langle X, \gamma a\beta\rangle \in \boldsymbol{\alpha}^{\ell}(I) \wedge \gamma \in \mathscr{T}^{\star}\}.\\
&= \mathscr{F}^{\sharp}_{\mathscr{G}} \circ \boldsymbol{\alpha}^{\ell}(I),
\end{aligned}$$

by defining

$$\begin{aligned}\mathscr{F}^{\sharp}_{\mathscr{G}}(R) \triangleq {} & \{\langle A, \alpha\rangle \mid A \xrightarrow{\mathscr{G}} \alpha\} \\ & \cup \{\langle Y, \beta\rangle \mid \langle X, \alpha Y \gamma\rangle \in R \wedge \alpha \in \mathscr{T}^{\star} \wedge Y \xrightarrow{\mathscr{G}} \beta\} \\ & \cup \{\langle X, \alpha\beta\gamma\rangle \mid \langle X, \alpha Y \gamma\rangle \in R \wedge \alpha \in \mathscr{T}^{\star} \wedge \langle Y, \beta\rangle \in R\} \\ & \cup \{\langle X, \alpha a\beta\rangle \mid \langle X, \alpha a \beta\rangle \in R \wedge \alpha \in \mathscr{T}^{\star}\}. \end{aligned} \tag{27}$$

By Lemma 2, we conclude that $\boldsymbol{\alpha}^{\ell}(\mathrm{lfp}^{\subseteq} \mathscr{F}_{\mathscr{G}}) = \mathrm{lfp}^{\subseteq} \mathscr{F}^{\sharp}_{\mathscr{G}}$. Since $\mathscr{F}^{\sharp}_{\mathscr{G}}(R) = \mathscr{D}^{\mathscr{G}}_{A,\ell}(R) \cup \{\langle X, \alpha a\beta\rangle \mid \langle X, \alpha a\beta\rangle {\in} R \wedge \alpha {\in} \mathscr{T}^{\star}\}$ and the last term $\{\langle X, \alpha a\beta\rangle \mid \langle X, \alpha a\beta\rangle {\in} R \wedge \alpha {\in} \mathscr{T}^{\star}\}$ of $\mathscr{F}^{\sharp}_{\mathscr{G}}(R)$ in (27) adds no new element to the transfinite iterates [5] of $\mathrm{lfp}^{\subseteq} \mathscr{F}^{\sharp}_{\mathscr{G}}(R)$, we have $\mathrm{lfp}^{\subseteq} \mathscr{F}^{\sharp}_{\mathscr{G}}(R) = \mathrm{lfp}^{\subseteq} \mathscr{D}^{\mathscr{G}}_{A,\ell}$ proving $\boldsymbol{\alpha}^{\ell}(\mathrm{lfp}^{\subseteq} \mathscr{F}_{\mathscr{G}}) = \mathrm{lfp}^{\subseteq} \mathscr{D}^{\mathscr{G}}_{A,\ell}$ whence (26) by (12) and (23). □

8. Item semantics-based specification of the language generated by a grammar

It directly follows from (26) that the language $\mathscr{L}_{\mathscr{G}}$ generated by a grammar $\mathscr{G}$ traditionally defined by (8) can be equivalently defined using the grammar item semantics $\mathscr{I}_{\mathscr{G}}$:

Corollary 5.

$$\mathscr{L}_{\mathscr{G}} = \{\gamma \in \mathscr{T}^{\star} \mid \exists \lambda \in \mathscr{T}^{\star}\colon\ [\lambda, A \to \alpha\cdot, \gamma] \in \mathscr{I}_{\mathscr{G}}\}. \tag{28}$$

Proof. We have

$$\begin{aligned} & \langle X, \delta\rangle \in \boldsymbol{\alpha}^{\ell}(\mathscr{I}_{\mathscr{G}}) \wedge \delta \in \mathscr{T}^{\star} \\ & \quad \Leftrightarrow \wr \text{by def. (24) of } \boldsymbol{\alpha}^{\ell} \wr \\ & \quad\ \exists \lambda, \beta, \gamma \in \mathscr{T}^{\star}\colon\ \delta = \gamma\beta \wedge [\lambda, X \to \alpha \cdot \beta, \gamma] \in \mathscr{I}_{\mathscr{G}} \\ & \quad \Rightarrow \wr \text{since } [\lambda, X \to \alpha \cdot \beta, \gamma] \in \mathscr{I}_{\mathscr{G}} \wedge \beta \in \mathscr{T}^{\star} \text{ implies } [\lambda, X \to \alpha\beta\cdot, \gamma\beta] \in \mathscr{I}_{\mathscr{G}} \\ & \quad\ \text{by (22) and induction on the length of } \beta \wr \\ & \quad\ \exists \lambda, \beta \in \mathscr{T}^{\star}\colon\ [\lambda, X \to \alpha\beta\cdot, \delta] \in \mathscr{I}_{\mathscr{G}} \\ & \quad \Rightarrow \wr \text{by def. (24) of } \boldsymbol{\alpha}^{\ell} \wr \\ & \quad\ \langle X, \delta\rangle \in \boldsymbol{\alpha}^{\ell}(\mathscr{I}_{\mathscr{G}}), \end{aligned}$$

proving that for $\delta {\in} \mathscr{T}^{\star}$ we have the equivalence

$$\langle X, \delta\rangle \in \boldsymbol{\alpha}^{\ell}(\mathscr{I}_{\mathscr{G}}) \quad \Leftrightarrow \quad \exists \lambda \in \mathscr{T}^{\star}\colon\ [\lambda, X \to \alpha\cdot, \delta] \in \mathscr{I}_{\mathscr{G}}. \tag{29}$$

We conclude that the language generated by the grammar $\mathscr{G}$ is

$$\begin{aligned}\mathscr{L}_{\mathscr{G}} &= \{\delta \in \mathscr{T}^\star \mid A \overset{\mathscr{G}}{\Rightarrow}_{A,\ell} \delta\} \quad \wr \text{by } (10) \wr\\ &= \{\delta \in \mathscr{T}^\star \mid \langle A, \delta\rangle \in \boldsymbol{\alpha}^{\ell}(\mathscr{I}_{\mathscr{G}})\} \quad \wr \text{by } (26) \wr\\ &= \{\delta \in \mathscr{T}^\star \mid \exists \lambda \in \mathscr{T}^\star\colon\ [\lambda, X \to \alpha\cdot, \delta] \in \mathscr{I}_{\mathscr{G}}\} \quad \wr \text{by } (29) \wr . \qquad \square\end{aligned}$$

9. Earley parsing algorithm is a complete abstraction of the grammar item semantics

The Earley's parsing algorithm (17) derives the only grammar items which are valid for the given input word $\omega = \omega_1 \ldots \omega_n$, $n \geqslant 0$ to be analyzed.

9.1. The abstraction

This is a forgetful abstraction disregarding all information provided by the grammar item semantics, but for the input word

$$\begin{aligned}\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(I) \triangleq \{\langle X \to \alpha \cdot \beta, i, j\rangle \mid 0 \leqslant i \leqslant j \leqslant n\\ \wedge [\omega_1 \ldots \omega_i, X \to \alpha \cdot \beta, \omega_{i+1} \ldots \omega_j] \in I\}.\end{aligned} \tag{30}$$

$\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}$ is a complete $\cup$-morphism so it is the upper adjoint of a Galois connection

$$\langle \wp(\mathscr{T}^\star \times \mathscr{N} \times \mathscr{V}^\star \times \mathscr{V}^\star \times \mathscr{T}^\star), \subseteq\rangle \underset{\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}}{\overset{\gamma^{\mathrm{E}}_{\omega}}{\leftrightarrows}} \langle \wp(\mathscr{N} \times \mathscr{V}^\star \times \mathscr{V}^\star \times \mathbb{N} \times \mathbb{N}), \subseteq\rangle. \tag{31}$$

9.2. The abstract interpretation of the semantics

By abstraction of the fixpoint definition (23) of the grammar item semantics with $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}$, we get the fixpoint characterization (17) of the Earley's valid item semantics

Theorem 6.

$$\mathscr{I}^{\mathrm{E}}_{\mathscr{G},\omega} = \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(\mathscr{I}_{\mathscr{G}}). \tag{32}$$

Proof. We must prove that $\mathscr{I}^{\mathrm{E}}_{\mathscr{G},\omega} = \mathrm{lfp}^{\subseteq} \mathscr{F}^{\mathrm{E}}_{\mathscr{G},\omega} = \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(\mathrm{lfp}^{\subseteq} \mathscr{F}_{\mathscr{G}}) = \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(\mathscr{I}_{\mathscr{G}})$ which, by Lemma 2, immediately follows from $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega} \circ \mathscr{F}_{\mathscr{G}} = \mathscr{F}^{\mathrm{E}}_{\mathscr{G},\omega} \circ \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}$. Because $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}$ is a complete $\cup$-morphism, it is sufficient to do prove that term by term. We have

- $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(\{[\varepsilon, A \to \cdot\beta, \varepsilon] \mid A \overset{\mathscr{G}}{\to} \beta\})$

$$\begin{aligned}&= \wr\text{Definition (30) of } \boldsymbol{\alpha}^{\mathrm{E}}_{\omega} \wr\\ &\quad \{\langle X \to \alpha \cdot \beta, i, j\rangle \mid 0 \leqslant i \leqslant j \leqslant n \wedge [\omega_1 \ldots \omega_i, X \to \alpha \cdot \beta, \omega_{i+1} \ldots \omega_j]\\ &\qquad = [\varepsilon, A \to \cdot\beta, \varepsilon] \wedge A \overset{\mathscr{G}}{\to} \beta\}\end{aligned}$$

$= \wr\, \omega_1 \ldots \omega_i = \varepsilon$ so $i = 0,\ X = A,\ \alpha \cdot \beta = \cdot\beta$ so $\alpha = \varepsilon$,

$\omega_{i+1} \ldots \omega_j = \omega_1 \ldots \omega_j = \varepsilon$ so $j = 0\, \wr$

$\{\langle A \to \cdot\gamma, 0, 0\rangle \mid A \xrightarrow{\mathscr{G}} \gamma\}$.

- $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(\{[\lambda\alpha, Y \to \cdot\delta, \varepsilon] \mid [\lambda, X \to \alpha \cdot Y\beta, \gamma] \in I \wedge Y \xrightarrow{\mathscr{G}} \delta\})$

  $= \wr$Definition (30) of $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}\, \wr$

  $\{\langle X \to \alpha \cdot \beta, i, j\rangle \mid 0 \leqslant i \leqslant j \leqslant n \wedge [\omega_1 \ldots \omega_i, X \to \alpha \cdot \beta, \omega_{i+1} \ldots \omega_j] \in$

  $\{[\lambda'\gamma, Y \to \cdot\delta, \varepsilon] \mid [\lambda', X \to \alpha' \cdot Y\beta', \gamma] \in I \wedge Y \xrightarrow{\mathscr{G}} \delta\}\}$

  $= \wr\, \lambda'\gamma = \omega_1 \ldots \omega_i$ so $\exists k \in [0, i]$: $\lambda' = \omega_1 \ldots \omega_k \wedge \gamma = \omega_{k+1} \ldots \omega_i,\ X = Y$,

  $\alpha \cdot \beta = \cdot\delta$ so $\alpha = \varepsilon$ and $\beta = \delta,\ \omega_{i+1} \ldots \omega_j = \varepsilon$ so $i = j\, \wr$

  $\{\langle Y \to \cdot\delta, j, j\rangle \mid 0 \leqslant k \leqslant j \leqslant n \wedge [\omega_1 \ldots \omega_k, X \to \alpha' \cdot Y\beta', \omega_{k+1} \ldots \omega_j] \in$

  $I \wedge Y \xrightarrow{\mathscr{G}} \delta\}$

  $= \wr\, \alpha \cdot \beta = \alpha' \cdot Y\beta'$ if and only if $\alpha = \alpha'$ and $\beta = Y\beta'$, renaming k as $i\, \wr$

  $\{\langle Y \to \cdot\delta, j, j\rangle \mid \langle X \to \alpha \cdot Y\beta, i, j\rangle \in \{\langle X \to \alpha' \cdot \beta', i, j\rangle \mid 0 \leqslant i \leqslant j \leqslant$

  $n \wedge [\omega_1 \ldots \omega_i, X \to \alpha' \cdot \beta', \omega_{i+1} \ldots \omega_j] \in I \wedge Y \xrightarrow{\mathscr{G}} \delta\}\}$

  $= \wr$Definition (30) of $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}\, \wr$

  $\{\langle Y \to \cdot\delta, j, j\rangle \mid \langle X \to \alpha \cdot Y\beta, i, j\rangle \in \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(I) \wedge Y \xrightarrow{\mathscr{G}} \delta\}$.

- $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(\{[\lambda, X \to \alpha Y \cdot \beta, \gamma\xi] \mid [\lambda, X \to \alpha \cdot Y\beta, \gamma] \in I \wedge [\lambda\gamma, Y \to \delta\cdot, \xi] \in I\})$

  $= \wr$Definition (30) of $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}\, \wr$

  $\{\langle X \to \alpha \cdot \beta, i, j\rangle \mid 0 \leqslant i \leqslant j \leqslant n \wedge [\omega_1 \ldots \omega_i, X \to \alpha \cdot \beta, \omega_{i+1} \ldots \omega_j] \in$

  $\{[\lambda, X \to \alpha' Y \cdot \beta', \gamma\xi] \mid [\lambda, X \to \alpha' \cdot Y\beta', \gamma] \in I \wedge [\lambda\gamma, Y \to \delta\cdot, \xi] \in I\}\}$

  $= \wr\, \lambda = \omega_1 \ldots \omega_i, \alpha = \alpha' Y,\ \beta = \beta',\ \gamma\xi = \omega_{i+1} \ldots \omega_j$

  so $\exists k \in [i, j]$: $\gamma = \omega_{i+1} \ldots \omega_k \wedge \xi = \omega_{k+1} \ldots \omega_j\, \wr$

  $\{\langle X \to \alpha' Y \cdot \beta', i, j\rangle \mid 0 \leqslant i \leqslant k \leqslant j \leqslant n \wedge [\omega_1 \ldots \omega_i, X \to \alpha' \cdot Y\beta',$

  $\omega_{i+1} \ldots \omega_k] \in I \wedge [\omega_1 \ldots \omega_i \omega_{i+1} \ldots \omega_k, Y \to \delta\cdot, \omega_{k+1} \ldots \omega_j] \in I\}$

$= \wr$Definition (30) of $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}\wr$

$$\{\langle X \to \alpha Y \cdot \beta, k, j\rangle \mid \langle X \to \alpha \cdot Y\beta, k, i\rangle \in \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(I) \wedge \langle Y \to \gamma\cdot, i, j\rangle \in \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(I)\}.$$

- $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(\{[\lambda, X \to \alpha a \cdot \beta, \gamma a] \mid [\lambda, X \to \alpha a \cdot \beta, \gamma a] \in I\})$

$= \wr$Definition (30) of $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}\wr$

$$\{\langle X \to \alpha \cdot \beta, i, j\rangle \mid 0 \leqslant i \leqslant j \leqslant n \wedge [\omega_1 \ldots \omega_i, X \to \alpha \cdot \beta, \omega_{i+1} \ldots \omega_j]$$

$$\in \{[\lambda, X \to \alpha' a \cdot \beta', \gamma a] \mid [\lambda, X \to \alpha' \cdot a\beta', \gamma] \in I\}\}$$

$= \wr \lambda = \omega_1 \ldots \omega_i, \alpha = \alpha' a, \beta = \beta'$ and $\gamma a = \omega_{i+1} \ldots \omega_j$

so $\gamma = \omega_{i+1} \ldots \omega_{j-1}$ and $a = \omega_j \wr$

$$\{\langle X \to \alpha' \omega_j \cdot \beta, i, j\rangle \mid 0 \leqslant i \leqslant j \leqslant n$$

$$\wedge [\omega_1 \ldots \omega_i, X \to \alpha' \cdot \omega_j \beta', \omega_{i+1} \ldots \omega_{j-1}] \in I\}$$

$= \wr$Definition (30) of $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}\wr$

$$\{\langle X \to \alpha \omega_j \cdot \beta, i, j\rangle \mid \langle X \to \alpha \cdot \omega_j \beta, i, j-1\rangle \in \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(I)\}. \qquad \square$$

10. Correctness of Earley's parsing algorithm

Earley's parsing algorithm approximates the grammar items for the given terminal input word. This word is in the language generated by the grammar only if it is recognized by a grammar item for the axiom, so

Corollary 7. *The Earley's parsing algorithm is correct in that* (18) *holds.*

Proof.

$\langle A \to \gamma\cdot, 0, n\rangle \in \mathscr{I}^{\mathrm{E}}_{\mathscr{G},\omega}$

$\Leftrightarrow \wr$by (32)$\wr$

$\langle A \to \gamma\cdot, 0, n\rangle \in \boldsymbol{\alpha}^{\mathrm{E}}_{\omega}(\mathscr{I}_{\mathscr{G}})$

$\Leftrightarrow \wr$by Definition (30) of $\boldsymbol{\alpha}^{\mathrm{E}}_{\omega}\wr$

$\langle A \to \gamma\cdot, 0, n\rangle \in \{\langle X \to \alpha \cdot \beta, i, j\rangle \mid 0 \leqslant i \leqslant j \leqslant n$

$\wedge [\omega_1 \ldots \omega_i, X \to \alpha \cdot \beta, \omega_{i+1} \ldots \omega_j] \in \mathscr{I}_{\mathscr{G}}\}$

$$\Leftrightarrow \wr X = A,\ \gamma\cdot = \alpha\cdot\beta \text{ so } \gamma = \alpha \text{ and } \beta = \varepsilon,\ i = 0,\ j = n$$

$$\text{so } \omega_1 \ldots \omega_i = \varepsilon \text{ and } \omega_{i+1} \ldots \omega_j = \omega \wr$$

$$[\varepsilon, A \to \alpha\cdot, \omega] \in \mathscr{I}_{\mathscr{G}}$$

$$\Leftrightarrow \wr \text{choosing } \lambda = \varepsilon \text{ and Lemma 3} \wr$$

$$\exists \lambda \in \mathscr{T}^{\star}\colon\ [\lambda, A \to \alpha\cdot, \omega] \in \mathscr{I}_{\mathscr{G}}$$

$$\Leftrightarrow \wr \text{by (28)} \wr$$

$$\omega \in \mathscr{L}_{\mathscr{G}}. \qquad \Box$$

11. Conclusion

We have shown that Earley's parsing algorithm [9] is an abstract interpretation of a refinement of the derivation semantics of grammars.

Other parsing algorithms may certainly be formally derived in a similar way using a more refined item semantics with nonterminal left and right contexts. A compile-time/static analysis of the grammar (item semantics) is used for top-down left-to-right generation of sets of grammar items abstracted e.g. as states. The same way a preliminary analysis of the grammar approximates terminal derivations from the right contexts by a lookahead. This preliminary static grammar analysis is used to ensure that the bottom-up recognition with left context is deterministic. This point of view remains to be applied, e.g. to LR-parsing [10].

References

[1] P. Aczel, An introduction to inductive definitions, in: J. Barwise (Ed.), Studies in Logic and the Foundations of Mathematics, Handbook of Mathematical Logic, Vol. 90, Elsevier, Amsterdam, 1977, pp. 739–782.

[2] P. Cousot, Types as abstract interpretations, invited paper, in: Conference Record of the 24th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, ACM Press, New York, 1997, pp. 316–331.

[3] P. Cousot, R. Cousot, Abstract interpretation of algebraic polynomial systems, in: M. Johnson (Ed.), Proc. 6th Internat. Conf. on Algebraic Methodology and Software Technology, AMAST '97, Sydney, Australia, Lecture Notes in Computer Science, Vol. 1349, Springer, Berlin, 1997, pp. 138–154.

[4] P. Cousot, R. Cousot, Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints, in: Conference Record of the Fourth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, ACM Press, New York, 1977, pp. 238–252.

[5] P. Cousot, R. Cousot, Constructive versions of Tarski's fixed point theorems, Pacific J. Math. 82 (1) (1979) 43–57.

[6] P. Cousot, R. Cousot, Systematic design of program analysis frameworks, in: Conf. Record of the Sixth Annual ACM SIGPLAN-SIGACT Symp. on Principles of Programming Languages, ACM Press, New York, 1979, pp. 269–282.

[7] P. Cousot, R. Cousot, Compositional and inductive semantic definitions in fixpoint, equational, constraint, closure-condition, rule-based and game-theoretic form, invited paper, in: P. Wolper (Ed.), Proc. 7th Internat. Conf. on Computer Aided Verification, CAV '95, Liège, Belgium, Lecture Notes in Computer Science, Vol. 939, Springer, Berlin, 1995, pp. 293–308.

[8] P. Cousot, R. Cousot, Temporal abstract interpretation, in: Conference Record of the 27th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, ACM Press, New York, 2000, pp. 12–25.

[9] J. Earley, An efficient context-free parsing algorithm, Communications of the Association for Computing Machinary 13 (2) (1970) 94–102.

[10] D.E. Knuth, On the translation of languages from left to right, Information and Control 8 (1965) 607–639.

[11] K. Sikkel, A. Nijholt, Parsing of context-free languages, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 2, Linear Modeling: Background and Application, Springer, Berlin, 1997, pp. 61–100.

[12] T.A. Sudkamp, Languages and Machines: An Introduction to the Theory of Computer Science, 2nd Edition, Addison-Wesley Pub. Co., Reading, MA, November 1996.

[13] A. Tarski, A lattice theoretical fixpoint theorem and its applications, Pacific J. Math. 5 (1955) 285–310.