

hal-00990007, version 1 - 13 May 2014

ISSN 0249-6399

ISRN INRIA/RR--8528--FR+ENG

A Grammatical Approach to Data-centric Case Management in a Distributed Collaborative Environment

Eric Badouel, Loïc Hélouët, Georges-Edouard Kouamou, Christophe Morvan

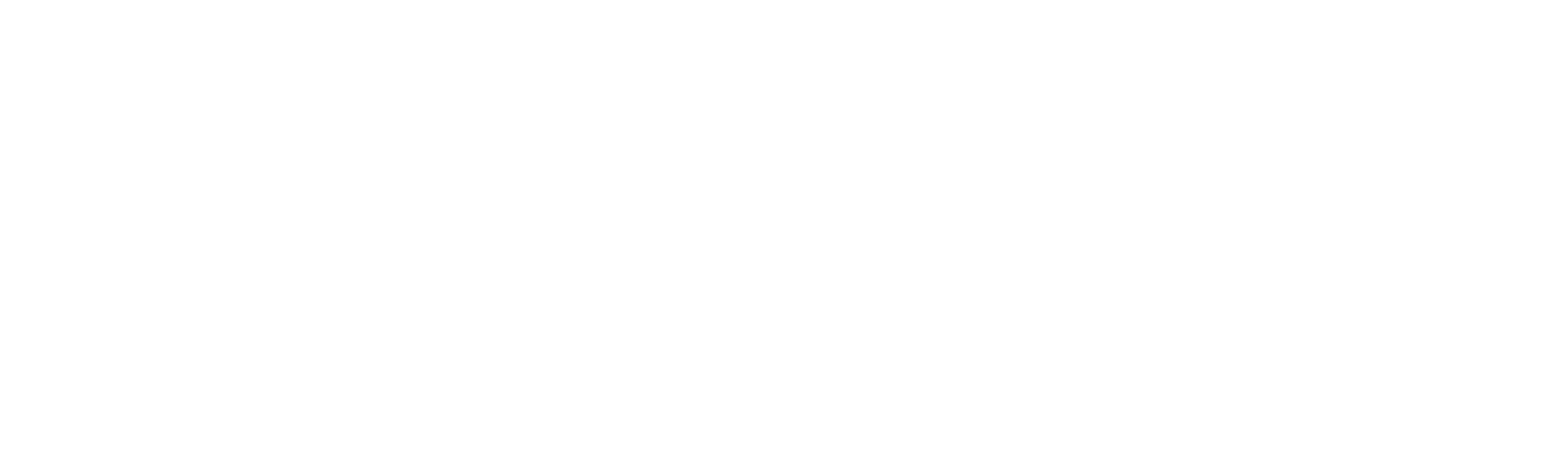
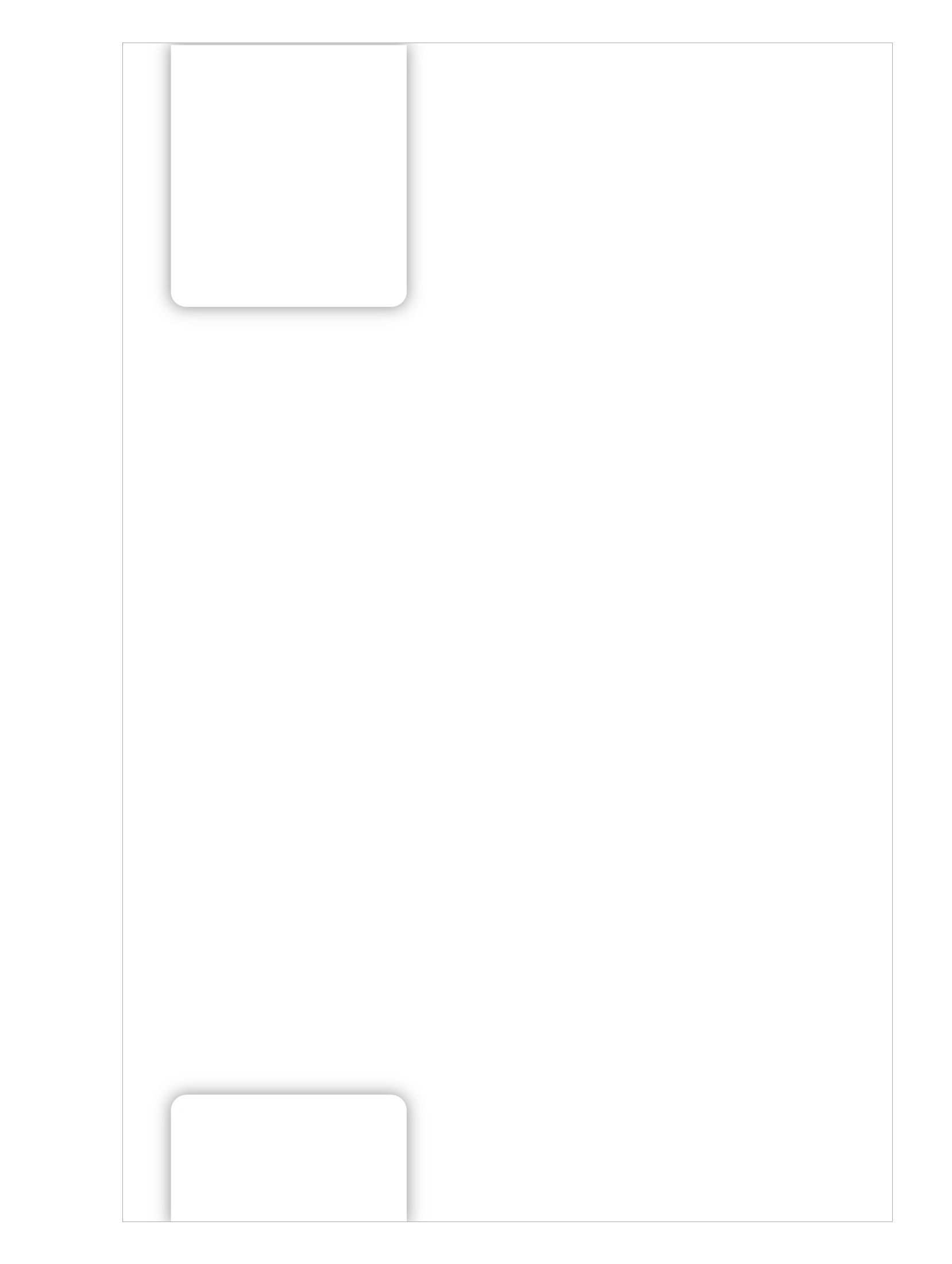
RESEARCH REPORT

N° 8528

May 2014

Project-Team SUMO

hal-00990007, version 1 - 13 May 2014



hal-00990007, version 1 - 13 May 2014

**A Grammatical Approach to Data-centric Case Management in a Distributed Collaborative Environment**

Eric Badouel∗, Loïc Hélouët†, Georges-Edouard Kouamou‡, Christophe Morvan §

Project-Team SUMO

Research Report n° 8528 — May 2014 — [32](#_bookmark24) pages

**Abstract:** This paper presents a purely declarative approach to artifact-centric case manage- ment systems, and a decentralization scheme for this model. Each case is presented as a tree-like structure; nodes bear information that combines data and computations. Each node belongs to a given stakeholder, and semantic rules govern the evolution of the tree structure, as well as how data values derive from information stemming from the context of the node. Stakeholders communicate through asynchronous message passing without shared memory, enabling convenient distribution.

**Key-words:** Active Documents, Business Artifacts, Case Management, Attribute Grammars

∗ Inria and Lirima

† Inria

‡ ENSP and Lirima

§ UniversitÃ© Paris-Est and Inria

**RESEARCH CENTRE**

**RENNES – BRETAGNE ATLANTIQUE**

Campus universitaire de Beaulieu 35042 Rennes Cedex

# Une approche grammaticale de la gestion de dossiers dans un environement collaboratif distribué

**Résumé :** Nous introduisons un modèle déclaratif de système de gestion de dossiers à base d’artefacts ainsi qu’un schéma de distribution de ce modèle. Chaque dossier a une structure arborescente dont les moeuds combinent données et calculs. Chaque noeud est sous la respon- sabilité d’un unique utilisateur et des règles sémantiques régissent les évolutions du document ainsi que le calcul de la valeur des attributs des noeuds. Les utilisateurs communiquent de façon asynchrone par échange de messages, et sans mémoire partagée, ce qui facilite la distribution du modèle.

**Mots-clés :** documents actifs, artefacts, gestion de dossiers, grammaires attribuées

hal-00990007, version 1 - 13 May 2014

*A Grammatical Approach to Data-centric Case Management* 3

# 1 Introduction

Traditional Case Management Systems rely on workflow models. The emphasis is put on the orchestration of activities involving humans (the *stakeholders*) and software systems, in order to achieve some global objective. In this context, stress is often put on control and coordination of the tasks required for the realization of a particular service. Such systems are usually modeled using centralized and state-based formalisms like automata, Petri nets or statecharts. They can also be directly specified with dedicated notations like BPEL [[22](#_bookmark22)] or BPMN 1.

A drawback of existing workflow formalisms is that *data* exchanged during the processing of a task play a secondary role when not simply ignored. However, data can be tightly connected with control flows and should not be overlooked. Actually, data contained in a request may influence its processing; conversely different decisions during the treatment of a case may produce distinct output-values.

Similarly, stakeholders are frequently considered as second class citizens in workflow systems: they are modeled as plain resources, performing specific tasks for a particular case, like machines in assembly lines. As a result, workflow systems are ideal to model fixed production schemes in manufactures or organizations, but can be too rigid to model open architectures where the evolving rules and data require more flexibility.

hal-00990007, version 1 - 13 May 2014

On the other hand, *data-centric workflow systems*, proposed by IBM [[21](#_bookmark22), [12](#_bookmark22), [5](#_bookmark21)], put stress on the exchanged documents, the so-called *Business Artifacts*, also known as *business entities with lifecycles*. An artifact is a document that conveys all the information concerning a particular case from its inception in the system until its completion. It contains all the relevant information about the entity together with a lifecycle that models its possible evolutions through the busi- ness process. Several variants presenting the life cycle of an artifact by an automaton, a Petri net [[17](#_bookmark22)], or logical formulas depicting legal successors of a state [[5](#_bookmark21)] have been proposed. However, even these variants remain state-based centralized models in which stakeholders are second class citizens.

Recently, Guard-Stage-Milestone (GSM), a declarative model of the lifecycle of artifacts has been introduced [[13](#_bookmark22), [6](#_bookmark21)]. This model defines *Guards*, *Stages* and *Milestones* to control the en- abling, enactment and completion of (possibly hierarchical) activities. The GSM lifecycle meta- model has been adopted as a basis of the OMG standard *Case Management Model and Notation* (CMMN). The GSM model allows for dynamic creation of subtasks (the *stages*), and handles data attributes. Furthermore, guards and milestones attached to stages provide declarative de- scriptions of tasks inception and termination. However, interaction with users are modeled as incoming messages from the environment, or as events from low-level (atomic) stages. In this way, users do not contribute to the choice of a workflow for a process. The semantics of GSM models is given in terms of global snapshots. Events can be handled by all stages as soon as they are produced, and guard of a stage can refer to attributes of distant stages. Thus this model is not directly executable on a distributed architecture.

*This paper introduces a distributed and declarative model for Case Management called Guarded Attribute Grammars (GAG for short), which is both data-centric, user-driven, and provides a convenient way to handle distribution*. GAGs are collections of semantic rules that describe how

to produce data from inputs provided by the environment. They are a variant of attribute gram- mars [[16](#_bookmark22), [23](#_bookmark22)]. Their notation is reminiscent of unification grammars, and is inspired by the work of Deransart and Maluszynski [[8](#_bookmark21)] relating attribute grammars with definite clause programs.

In this declarative model, the lifecycle of artifacts is left implicit. Artifacts under evaluation can be seen as incomplete structured documents, i.e., trees with *open nodes* corresponding to parts of the document that remain to be completed. Each open node is attached intentional data,



1 [www.bpmn.org](http://www.bpmn.org/)

RR n° 8528

4 *Badouel & al.*

i.e., an expression for the piece of information to be substituted to the node. The evolution of an artifact is governed by the stakeholder’s decisions (choosing a particular action amongst those which are enabled at a given moment, inputing data,...), and by application of the semantic rules to update artifacts (by refining open nodes). Cases reach completion when they do not contain open nodes. An artifact is thus a structured document with some active parts. This notion of *active documents* is close to the model of Active XML introduced by Abiteboul et al. [[1](#_bookmark21)] which consists of semi-structured documents with embedded service calls.

The paper is organized as follows. After an informal introduction to our grammatical ap- proach to case management in Section [2](#_bookmark0) we introduce the model of *guarded attribute grammars* that underpins the approach by presenting successively its syntax (Section [3](#_bookmark3)), its behaviour (Section [4](#_bookmark6)) and by giving some illustrative examples (Section [5](#_bookmark11)). The deployment of a guarded attribute grammar on a distributed asynchronous architecture is studied in Section [6](#_bookmark16). Soundness of guarded attribute grammars is touched upon in Section [7](#_bookmark20). An assessment of the model and future research directions are given in conclusion.

**2 A Grammatical Approach to Case Management**

hal-00990007, version 1 - 13 May 2014

This section introduces a grammatical notation for case management which will be described more formally in the subsequent two sections.

Throughout the paper, the term *case* designates a concrete instance of a given business process. We use the editorial process of an academic journal as a running example to illustrate

the various notions and notations. A case for this example is the editorial processing of a particular article submitted to the journal.

The case is handled by various actors involved in the process, called *stakeholders*, namely the editor in chief, an associate editor and some referees. We associate each case with a document, called an *artifact*, that collects all the information related to the case from its inception in the process until its completion. When the case is closed this document constitutes a full history of all the decisions that led to its completion.

We interpret a case as a problem to be solved, that can be completed by refining it into sub-tasks using business rules. This notion of business rule can be modelled by a *production P* : *s*0 ← *s*1 · · · *sn* expressing that task *s*0 can be reduced to subtasks *s*1 to *sn*. If several productions with the same left-hand side *s*0 exist then the choice of a particular production corresponds to a decision made by some designated stakeholder. For instance, there are two possible immediate outcomes for a submitted article: either it is validated by the editor in chief

and it enters the evaluation process of the journal or it is invalidated because its topic or format is not adequate. This initial decision can be reflected by the two following productions:

validate : Proposed\_submission ← Submission

invalidate : Proposed\_submission ←

If *P* is the unique production having *s*0 in its left-hand side, then there is no real decision to make and such a rule is interpreted as a logical decomposition of the task *s*0 into substasks *s*1 to *sn*. Such a production will be automatically triggered without human intervention.

Accordingly, we model an artifact as a tree whose nodes are sorted. We write *X* :: *s* to indicate

that node *X* is of sort *s*. An artifact is given by a set of equations of the form *X* = *P* (*X*1*, . . . , Xn*), stating that *X* :: *s* is a node labeled by production *P* : *s* ← *s*1 · · · *sn* and with successor nodes *X*1 :: *s*1 to *Xn* :: *sn*. In that case node *X* is said to be a *closed* node defined by equation *X* = *P* (*X*1*, . . . , Xn*) (we henceforth assume that we do not have two equations with the same left-hand side). A node *X* :: *s* defined by no equation (i.e. that appears only in the right hand side of an equation) is an *open node*. It corresponds to a pending task *s*.

Inria

*A Grammatical Approach to Data-centric Case Management* 5

The lifecycle of an artifact is implicitly given by a set of productions:

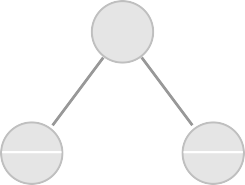
1. The artifact initially associated with a case is reduced to a single open node.
2. An open node *X* of sort *s* can be *refined* by choosing a production

*P* : *s* ← *s*1 *. . . sn* that fits its sort.

*s s*

The open node *X* becomes a closed node *X* =

?



*P*

*s*1

*sn*

?

?

*P* (*X*1*, . . . , Xn*) under the decision of applying production *P* to it. In doing so the task *s* associated with *X* is replaced by *n* subtasks *s*1 to *sn* and new open nodes *X*1 :: *s*1 to *Xn* :: *sn* are created accordingly.

1. The case has reached completion when its associated artifact is closed, i.e. it no longer contains open nodes.

However, plain context-free grammars do not model the interactions and data exchanged between the various tasks associated with open nodes. To overcome this problem, we attach additional information to open nodes using *attributes*. Each sort *s* ∈ *S* comes equipped with a set of *inherited* attributes and a set of *synthesized* attributes. Values of attributes are given by

hal-00990007, version 1 - 13 May 2014

*terms* over a ranked alphabet. Recall that such a term is either a variable or an expression of the form *c*(*t*1*, . . . , tn*) where *c* is a symbol of rank *n*, and *t*1*, . . . , tn* are terms. In particular a constant *c*, i.e. a symbol of rank 0, will be identified with the term *c*( ). We will denote by *var*(*t*) the set of variables used in term *t*.

**Definition 2.1** (Forms)**.** *A* **form** *of sort s is an expression*

*F* = *s*(*t*1*, . . . , tn*)(*u*1*, . . . , um*)

*where t*1*, . . . , tn (respectively u*1*, . . . , um) are terms over a ranked alphabet —the alphabet of attribute’s values— and a set of variables var* (*F* )*. Terms t*1*, . . . , tn give the values of the* **inherited attributes** *and u*1*, . . . , um the values of the* **synthesized attributes***) attached to form F.* ✷ *Def.* [*2.1*](#_bookmark1)

From now on, we consider productions where sorts are replaced by forms of the corresponding sorts. More precisely, a production is of the form

*s*0(*p*1*, . . . , pn*)(*u*1*, . . . , um*) ← *s*1(*t*(1)*, . . . , t*(1))(*y*(1)*, . . . , y*(1))

1

· · ·

*n*1 1 *m*1

(1)

*sk*(*t*(*k*)

(*k*)

(*k*)

(*k*)

1 *, . . . , tnk* )(*y*1 *, . . . , ymk* )

where the *pi*’s, the *uj* ’s, and the *t*(*ℓ*)’s are terms and the *y*(*ℓ*)’s are variables. The forms in the right-

*j j*

hand side of a production are *service calls*, namely they are forms *F* = *s*(*t*1*, . . . , tn*)(*y*1*, . . . , ym*)

where the synthesized positions are (distinct) variables *y*1*, . . . , ym* (i.e., they are not instantiated). The rationale is that we invoke a service by filling in the inherited positions of the form (the entries) and by indicating the variables that expect to receive the results returned by the service (the subscriptions).

Any open node is now attached to a service call. The corresponding service is supposed to *(i)* construct the tree that will refine the open node and *(ii)* compute the values of the synthesized attributes (i.e., it should return the subscribed values). A service is enacted by

RR n° 8528

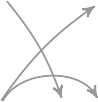
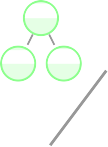
6 *Badouel & al.*

applying productions. More precisely, a production such as the one given in formula ([1](#_bookmark1)) can apply in an open node *X* when its left-hand side matches with the service call *s*0(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) attached to node *X*. For that purpose the terms *pi*’s are used as patterns that should match the corresponding data *di*’s. When the production applies, new open nodes are created and they are respectively associated with the forms (service calls) in the right-hand side of the production. The values of *uj* ’s are then returned to the corresponding variables *yj* ’s that had subscribed to these values. For instance applying production

*P* : *s*0(*a*(*x*1*, x*2))(*b*(*y*′ )*, y*′ ) ← *s*1(*c*(*x*1))(*y*′ ) *s*2(*x*2*, y*′ )(*y*′ )

1 2 1 1 2

to a node associated with service call *s*0(*a*(*t*1*, t*2))(*y*1*, y*2) gives rise to the substitution *x*1 = *t*1 and *x*2 = *t*2. The two newly-created open nodes are respectively associated



*a*

? ?

*x*1 *x*2 *s*1

*s*0

P

*b*

*y*′ *y*′

1 2

*s*2

?

?

*c*

?

*x*2

*y*′

1

?

′

*x*1

*y*1

*y*2

′

with the service calls *s*1(*c*(*t*1))(*y*′ ) and *s*2(*t*2*, y*′ )(*y*′ ) and

1 1 2

the values *b*(*y*′ ) and *y*′ are substituted to the variables *y*1

1 2

and *y*2 respectively.

hal-00990007, version 1 - 13 May 2014

The precise definitions are given in the next section. For the moment, let us illustrate the notations on our running example. A stakeholder has a specific **role** in the editorial process: he can be an author, the editor in chief, an associate editor or a referee. Each role is associated with a set of services and a set of productions explaining how each service is provided. For

instance an associate editor provides the service Submission(*article* )(*decision* ) consisting in returning an editorial decision about an article submitted to the journal. We emphasize the fact that production MakeDecision(*decision* ) : Decide(*report* 1*, report*2)(*decision* ) ← has a parameter

*decision*, that is used to enter new data in the case. Parameters are a convenient way to avoid

specifing a production for each allowed parameter. A parametric production can be equivalently replaced by several non-parametric productions when its parameters range over a finite set of values. The corresponding productions are listed in Table [1](#_bookmark2).

Table 1: Acting as an associate Editor

DecideSubmission : Submission(*article* )(*decision* ) ←

Evaluate(*article* )(*report*1)

Evaluate(*article* )(*report*2)

Decide(*report*1*, report*2)(*decision* )

MakeDecision(*decision* ) : Decide(*report*1*, report*2)(*decision* ) ←

AskReview(*reviewer* ) : Evaluate(*article* )(*report*) ←

WaitReport(*answer, article* )(*report* )

Call(*reviewer,* ToReview(*article* )(*answer* )) CaseNo(*msg* ) : WaitReport(No(*msg* )*, article* )(*report* ) ←

Evaluate(*article* )(*report*)

CaseYes(*msg* ) : WaitReport(Yes(*msg, report*)*, article* )(*report*) ←

The first two productions mean that an associate editor makes an editorial decision about a submitted paper on the basis of the evaluation reports produced by two different referees. He can

ask a report from a reviewer through an invocation of the external service ToReview(*article* )(*answer* ). The productions that govern the actions of a reviewer are given in Table [2](#_bookmark3).

One can group the productions of Table [1](#_bookmark2) and Table [2](#_bookmark3) using an additional parameter *reviewer*

to make as many disjoint copies of the specification given in Table [2](#_bookmark3) as there are individuals

Inria

*A Grammatical Approach to Data-centric Case Management* 7

Table 2: Acting as a reviewer

Decline(*msg* ) : ToReview(*article* )(No(*msg* )) ←

Accept(*msg* ) : ToReview(*article* )(Yes(*msg, report*)) ← Review(*article* )(*report* )

MakeReview(*report*) : Review(*article* )(*report* ) ←

playing the role of a referee. The resulting set of productions (where call to external services have been eliminated) is given in Table [3](#_bookmark3). Similarly one has as many instances of the productions in

Table 3: Making a decision on a submitted paper

DecideSubmission : Submission(*article* )(*decision* ) ←

Evaluate(*article* )(*report* 1)

Evaluate(*article* )(*report* 2)

Decide(*report* 1*, report*2)(*decision* )

MakeDecision(*decision* ) : Decide(*report* 1*, report*2)(*decision* ) ←

hal-00990007, version 1 - 13 May 2014

AskReview(*reviewer* ) : Evaluate(*article* )(*report* ) ←

WaitReport(*answer, article* )(*report*)

ToReview(*reviewer , article* )(*answer* ) Decline(*msg* )(*reviewer* ) : ToReview(*reviewer , article* )(No(*msg* )) ← Accept(*msg* )(*reviewer* ) : ToReview(*reviewer , article* )(Yes(*msg, report*)) ←

Review(*reviewer, article* )(*report*)

MakeReview(*report* )(*reviewer* ) : Review(*reviewer, article* )(*report*) ←

CaseNo(*msg* ) : WaitReport(No(*msg* )*, article* )(*report* ) ←

Evaluate(*article* )(*report* )

CaseYes(*msg* ) : WaitReport(Yes(*msg, report*)*, article* )(*report* ) ←

Table [1](#_bookmark2) as there are associate editors in the editorial board. In the complete (flat) specification one should therefore add an additional parameter *associateEditor* to distinguish between all associate editors. If the specification is large and contains many different roles the resulting global grammar can be quite complex. Yet, it is still possible to build an equivalent monolithic grammar without external service calls.

The above specification uses production schemes rather than plain productions. Therefore the actual productions of the grammar are instances of these productions schemes where specific values are substituted to the parameters. Replacing all parameters by their possible values to obtain plain productions in a systematic way results in a *guarded attribute grammar* (defined in Section [3](#_bookmark3)) with an infinite set of productions. However, at least in the above example, the parameters of the productions correspond either to a specific role in the process or to some kind of data (a message, a report, a decision) whose precise value has no impact on the behavior of the system. Therefore one can abstract this specification by identifying all individuals playing the same role and by representing each type of data by a corresponding constant so that one can obtain a finite guarded attribute grammar with the same behavior.

# 3 The Syntax of Guarded Attribute Grammars

Attribute grammars, introduced by Donald Knuth in the late sixties [[16](#_bookmark22)], have been instrumental in the development of syntax-directed transformations and compiler design. More recently this

RR n° 8528

8 *Badouel & al.*

model has been revived for the specification of structured document’s manipulations mainly in the context of web-based applications. The expression *grammareware* has been coined in [[15](#_bookmark22)] to qualify the tools for the design and customization of grammars and grammar-dependent softwares. One such interesting tool is the UUAG system developped by Swierstra and his group. They relied on purely functional implementations of attribute grammars [[14](#_bookmark22), [25](#_bookmark22), [2](#_bookmark21)] to build a domain specific languages (DSL) as a set of functional combinators derived from the semantic rules of an attribute grammar [[26](#_bookmark23), [25](#_bookmark22), [24](#_bookmark22)]. We intend to adapt this construction to the model of guarded attribute grammars introduced in this paper.

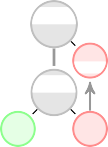
An Attribute grammar is obtained from an underlying grammar by associating each sort *s* with a set *Att*(*s*) of *attributes* —which henceforth should exist for each node of the given sort— and by associating each production *P* : *s* ← *s*1 *. . . sn* with semantic rules describing the functional dependencies between the attributes of a node labelled *P* (hence of sort *s*) and the attributes of its successor nodes (of respective sorts *s*1 to *sn*).

We use a non-standard notation for attribute grammars, inspired from [[7](#_bookmark21), [8](#_bookmark21)]. Let us introduce this notation on an example before proceeding to the formal definitions.

**Example 3.1** (Flattening of a binary tree)**.** *Our first illustration is the classical example of the attribute grammar that computes the flattening of a binary tree, i.e., the sequence of the leaves read from left to right. The semantic rules are usually presented as shown in Table* [*4*](#_bookmark4)*. The sort bin*

hal-00990007, version 1 - 13 May 2014

*X* :: *root*



Root

*x*

?

Nil

?

Table 4: Flattening of a binary tree

Root: (*X* :: *root*) ← (*X*1 :: *bin* )

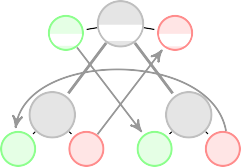
Where X· *s =* *X*1 · *s*

*X*1 · *h* = Nil

*x*

*X* :: *bin*

Fork : (*X* :: *bin* ) ← (*X*1 :: *bin* ) (*X*2 :: *bin*)



*x*

Fork

?

*y*

?

?

*z*

?

*x*

?

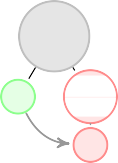
Where X· *s =* *X*1 · *s*

*X*1 · *h* = *X*2 · *s X*2 · *h* = *X* · *h*

*y z*

*X* :: *bin*

Leaf*a* : (*X* :: *bin*) ←



Leaf*a*

?

Cons*a*

*x*

*x* where *X* · *s* = Cons*a*(*X* · *h*)

*of binary trees has two attributes: the inherited attribute h contains an accumulating parameter and the synthesized attribute s eventually contains the list of leaves of the tree appended to the accumulating parameter. Which we may write as t* · *s* = *flatten*(*t*) ++*t* · *h, i.e., t* · *s* = *flat*(*t, t* · *h*)

*where flat*(*t, h*) = *flatten*(*t*) ++*h. The semantics rules stem from the identities:*

*flatten*(*t*)=*flat*(*t, Nil*)

*flat*(Fork(*t*1*, t*2)*, h*)=*flat*(*t*1*, flat*(*t*2*, h*))

*flat*(Leaf*a, h*)=Cons*a*(*h*)

Inria

*A Grammatical Approach to Data-centric Case Management* 9

*We present the semantics rules of Table* [*4*](#_bookmark4) *using the following syntax:*

Root : *root*()(*x*) ← *bin*(*Nil*)(*x*)

Fork : *bin* (*x*)(*y*) ← *bin*(*z*)(*y*) *bin* (*x*)(*z*)

Leaf*a* : *bin*(*x*)(Cons*a*(*x*)) ←

*The syntactic categories of the grammar, also called its sorts, namely root and bin are associated with their inherited attributes (given as a list of arguments:* (*t*1*, . . . , tn*)*) and their synthesized attributes (the co-arguments:*(*u*1*, . . . , um*)*). A variable x is an input variable, denoted as x*?*, if it appears in an inherited attribute of the left-hand side or in a synthesized attribute of the right-*

*hand side. It corresponds to a piece of information stemming respectively from the context of the node or from the subtree rooted at the corresponding successor node. These variables should be pairwise distinct. Symmetrically a variable is an output variable, denoted as x*! *, if it appears in*

*a synthesized attribute of the left-hand side or in an inherited attribute of the right-hand side. It corresponds to values computed by the semantic rules and send respectively to the context of the node or the subtree rooted at the corresponding successor node. Indeed, if we annotate the occurrences of variables with their polarity (input or output) one obtains:*

hal-00990007, version 1 - 13 May 2014

Root : *root*()(*x*!) ← *bin* (*Nil*)(*x*?)

Fork : *bin* (*x*?)(*y*!) ← *bin* (*z*!)(*y*?) *bin*(*x*!)(*z*?)

Leaf*a* : *bin* (*x*?)(Cons*a*(*x*!)) ←

*And if we draw an arrow from the (unique) occurrence of x*? *to the (various) occurrences of x*! *for each variable x to witness the data dependencies then the above rules correspond precisely to the three figures shown on the left-hand side of Table* [*4*](#_bookmark4)*.* ✷ *Exple* [*3.1*](#_bookmark4)

Guarded attribute grammars extend the traditional model of attribute grammars by allowing patterns rather that plain variables (as it was the case in the above example) to represent the inherited attributes in the left-hand side of a production. Patterns allow the semantic rules to process by case analysis based on the shape of some of the inherited attributes, and in this way to handle the interplay between the data (contained in the inherited attributes) and the control (the enabling of productions).

**Definition 3.2** (Guarded Attribute Grammars)**.** *Given a set of sorts S with fixed inherited and synthesized attributes. A* ***guarded attribute grammar*** *is a set of productions P* : *F*0 ← *F*1 · · · *Fk where the Fi* :: *si are forms. The inherited attributes of left-hand side F*0 *are called the* ***patterns*** *of the production. The values of synthesized attributes in the right-hand side are*

*variables. These occurrence of variables together with the variables occurring in the patterns are called the* **input occurrences** *of variables. We assume that each variable has* **at most one input occurrence***.* ✷ *Def.* [*3.2*](#_bookmark5)

The well-formedness conditions of GAGs express that every output is defined in terms of the inputs. We will often refer to this correspondences as the **semantic rules**. More precisely, the inputs are associated with (distinct) variables and the value of each output is given by a term using these variables.

Each variable can have several occurrences. First it should appear once as an input and it may also appear in several occurrences within some output term. The corresponding occurrence is respectively said to be in an *input* or in an *output position*. One can define the following transformation on productions whose effect is to annotate each occurrence of a variable so that *x*? (respectively *x*!) stands for an occurrence of *x* in an input position (resp. in an output

RR n° 8528

10 *Badouel & al.*

position).

!(*F*0 ← *F*1 · · · *Fk*) = ?(*F*0) ←!(*F*1) · · ·!(*Fk* )

?(*s*(*t*1*, . . . tn*)(*u*1*, . . . um*)) = *s*(?(*t*1)*, . . .*?(*tn*))(!(*u*1)*, . . .*!(*um*))

!(*s*(*t*1*, . . . tn*)(*u*1*, . . . um*)) = *s*(!(*t*1)*, . . .*!(*tn*))(?(*u*1)*, . . .*?(*um*))

?(*c*(*t*1*, . . . tn*)) = *c*(?(*t*1)*, . . .*?(*tn*))

!(*c*(*t*1*, . . . tn*)) = *c*(!(*t*1)*, . . .*!(*tn*))

?(*x*) = *x*?

!(*x*) = *x*!

The conditions stated in Definition [3.2](#_bookmark5) say that in the labelled version of a production each vari- able occurs at most once in an input position, i.e., that {?(*F*0)*,* !(*F*1)*, . . . ,* !(*Fk* )} is an admissible labelling of the set of forms in *P* according to the following definition.

**Definition 3.3** (Link Graph)**.** *A labelling in* {?*,* !} *of the variables var* (F) *of a set of forms* F *is* ***admissible*** *if the labelled version of a form F* ∈ F *is given by either* !*F or* ?*F and each variable has at most one occurrence labelled with* ?*. The occurrence x*? *identifies the place where the value of variable x is defined and the occurrences of x*! *identify the places where this value is used. The* ***link graph*** *associated with an admissible labelling of a set of forms* F *is the directed graph whose vertices are the occurrences of variables with an arc from v*1 *to v*2 *if these vertices are occurrences of a same variable x, labelled* ? *in v*1 *and* ! *in v*2*. This arc, depicted as follows,*

hal-00990007, version 1 - 13 May 2014

*x*



?

*x*

*means that the value produced in the* ***source vertex*** *v*1 *should be forwarded to the* ***target vertex***

*v*2*. Such an arc is called a* ***data link****.* ✷ *Def.* [*3.3*](#_bookmark6)

# 4 The Behaviour of Guarded Attribute Grammars

Attribute grammars are applied to input abstract syntax trees. These trees are usually produced by some parsing algorithm during a previous stage. The semantic rules are then used to decorate the node of the input tree by attribute values. In our setting the generation of the tree and its evaluation using the semantic rules are intertwined since the input tree represents an artifact under construction. An artifact is thus an incomplete abstract syntax tree which contains closed and open nodes. A closed node is labelled by the production that was used to create it. An open node is associated with a form that contains all the needed information for its further refinements. The information attached to an open node consists of the sort of the node and the current value of its attributes. The synthesized attributes of an open node are undefined and are thus associated with variables.

**Definition 4.1** (Configuration of a Guarded Attribute grammar)**.** *A* **configuration** Γ *of a guarded attribute grammar is an S-sorted set of nodes X* ∈ *nodes* (Γ) *each of which is associated with a defining equation in one of the following form where var* (Γ) *is a set of variables associated with* Γ*:*

**Closed node:** *X* = *P* (*X*1*, . . . , Xk*) *where P* : *s* ← *s*1 *. . . sk is a production of the underlying grammar and X* :: *s, and Xi* :: *si for* 1 ≤ *i* ≤ *k. Production P is the* **label** *of node X and nodes X*1 *to Xn are its* **successor nodes***.*

**Open node:** *X* = *s*(*t*1*, . . . , tn*)(*x*1*, . . . , xm*) *where X is of sort s and t*1*, . . . , tk are terms with variables in var* (Γ) *that represents the values of the inherited attributes of X, and x*1*, . . . , xm are variables in var* (Γ) *associated with its synthesized attributes.*

Inria

*A Grammatical Approach to Data-centric Case Management* 11

*Each variable in var* (Γ) *occurs at most once in a synthesized position. Otherwise stated* !Γ =

{!*F* | *F* ∈ Γ } *is an admissible labelling of the set of forms occurring in* Γ*.* ✷ *Def.* [*4.1*](#_bookmark6)

In order to specify the effect of applying a production at a given node of a configuration (Definition [4.4](#_bookmark8)) we first recall some notions about substitutions.

**Recall 4.2** (on Substitutions)**.** *We identify a substitution σ on a set of variables* {*x*1*, . . . , xk*}*, called the* ***domain*** *of σ, with a system of equations*

{*xi* = *σ*(*xi*) | 1 ≤ *i* ≤ *k* }

*The set of variables of σ, defined by* var(*σ*) = S1≤*i*≤*k* var(*σ*(*xi*))*, is disjoint from the domain of σ. Conversely a system of equations* {*xi* = *ti* | 1 ≤ *i* ≤ *k* } *defines a substitution σ with σ*(*xi*) = *ti if it is in* ***solved form****, i.e., none of the variables xi appears in some of the terms tj. In order to transform a system of equations E* = {*xi* = *ti* | 1 ≤ *i* ≤ *k* } *into an equivalent system*

.*xi* = *t*′ | 1 ≤ *j* ≤ *m* . *in solved form one can iteratively replace an occurrence of a variable xi in*

*j*

*one of the right-hand side term tj by its definition ti until no variable xi occurs in some tj. This process terminates when the relation xi* ≻ *xj* ⇔ *xj* ∈ var(*σ*(*xi*)) *is acyclic. One can easily verify that, under this assumption, the resulting system of equation SF* (*E*) = {*xi* = *t*′ | 1 ≤ *i* ≤ *n* } *in solved does not depend on the order in which the variables xi have been eliminated from the*

*i*

hal-00990007, version 1 - 13 May 2014

*right-hand sides. When the above condition is met we say that the set of equations is* ***acyclic***

*and that it* ***defines*** *the substitution associated with its solved form.* ✷ *Recall* [*4.2*](#_bookmark7)

The composition of two substitutions *σ, σ*′ is denoted by *σσ*′ and defined by *σσ*′ = {*x* = *tσ*′|*x* = *t* ∈ *σ*}. Similarly, we let Γ*σ* denote the configuration obtained from Γ by replacing the defining equation *X* = *F* of each open node *X* by *X* = *Fσ*.

We now define more precisely when a production is enabled at a given open node of a config- uration and the effect of applying the production. First note that variables of a production are formal parameters which scope is limited to the production. They can injectively be renamed in order to avoid clashes with variables names appearing in a configuration. Therefore we shall always assume that the set of variables of a production *P* is disjoint from the set of variables of a configuration Γ when applying production *P* at a node of Γ. As informally stated in the previous section, a production *P* applies at an open node *X* when its left-hand side *s*(*p*1*, . . . , pn*)(*u*1*, . . . um*)

matches with the definition *X* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*), i.e., the service call attached to *X* in

Γ.

First, the patterns *pi* should match with the data *di* according to the usual pattern matching given by the following inductive statements

match(*c*(*p*′ *, . . . , p*′ )*, c*′(*d*′ *, . . . , d*′ )) with *c* ƒ= *c*′ fails

1 *k* 1 *k*′

match(*c*(*p*′ *, . . . , p*′ )*, c*(*d*′ *, . . . , d*′ )) = .*k*

match(*p*′ *, d*′ )

1 *k* 1 *k i*=1 *i i*

match(*x, d*) = {*x* = *d*}

where the sum *σ* = .*k*

*i*=1

*σi* of substitutions *σi* is defined and equal to S

*i*∈1*..k*

*σi* when all

substitutions *σi* are defined and associated with disjoint sets of variables. Note that since no

variable occurs twice in the whole set of patterns *pi*, the various substitutions match(*pi, di*), when defined, are indeed concerned with disjoint sets of variables. Note also that match(*c*()*, c*()) = ∅.

**Definition 4.3.** *A form F* = *s*(*p*1*, . . . , pn*)(*u*1*, . . . um*) **matches** *with a service call*

*F* ′ = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) *(of the same sort) when*

1. *the patterns pi’s matches with the data di’s, defining a substitution σin* = .1≤*i*≤*n* match(*ti, di*)*,*

RR n° 8528

12 *Badouel & al.*

1. *the set of equations* {*yj* = *ujσin* | 1 ≤ *j* ≤ *m* } *is acyclic and defines a substitution σout . The resulting substitution σ* = match(*F, F* ′) *is given by σ* = *σout* ∪ *σin σout .* ✷ *Def.* [*4.3*](#_bookmark7)

**Definition 4.4** (Applying a Production)**.** *Let P* = *F* ← *F*1 *. . . Fk be a production,* Γ *be a configuration, and X be an open node with definition X* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) *in* Γ*. We assume that P and* Γ *are defined over disjoint sets of variables. We say that P is* ***enabled*** *in X and write* Γ[*P/X*)*, if the left-hand side of P matches with the definition of X. Then applying*

*production P at node X transforms configuration* Γ *into* Γ′*, denoted as* Γ[*P/X*)Γ′*, where:*

Γ′ = {*X* = *P* (*X*1*, . . . , Xk*)} *where X*1*, . . . , Xk are new nodes added to* Γ′

∪ {*X*1 = *F*1*σ, . . . , Xk* = *Fkσ*}

∪ {*X*′ = *Fσ* | (*X*′ = *F* ) ∈ Γ ∧ *X*′ ƒ= *X* }

*where σ* = match(*F, X*)*.* ✷ *Def.* [*4.4*](#_bookmark8)

Thus the first effect of applying production *P* to an open node *X* is that *X* becomes a closed node with label *P* and successor nodes *X*1 to *Xk*. The latter are new nodes added to Γ′. They are associated respectively with the instances of the *k* forms in the right-hand side of *P* obtained by applying substitution *σ* to these forms. The definitions of the other nodes of Γ are updated using substitution *σ* (or equivalently *σout* ). This update has no effect on the closed nodes because their defining equations in *Γ* contain no variable.

hal-00990007, version 1 - 13 May 2014

We conclude this section with two results justifying Definition [4.4](#_bookmark8). Namely, Proposition [4.5](#_bookmark8) states that if *P* is a production enabled in a node *X*0 of a configuration Γ with Γ[*P/X*0)Γ′ then Γ′ is a configuration, i.e., applying *P* cannot create a variable with several input occurrences. And Proposition [4.7](#_bookmark9) shows that the substitution *σ* = match(*F, X*) resulting from the matching of the left-hand side *F* = *s*(*p*1*, . . . , pn*)(*u*1*, . . . , um*) of a production *P* with the definition *X* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) of an open node *X* is the most general unifier of the set of equations

{*pi* = *di* | 1 ≤ *i* ≤ *n* } ∪ {*yj* = *uj* | 1 ≤ *j* ≤ *m* }.

**Proposition 4.5.** *If production P is enabled in an open node X*0 *of a configuration* Γ *and*

Γ[*P/X*0)Γ′ *then* Γ′ *is a configuration.*

*Proof.* Let *P* = *F* ← *F*1 *. . . Fk* with left-hand side *F* = *s*(*p*1*, . . . , pn*)(*u*1*, . . . um*) and *X*0 = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) be the defining equation of *X*0 in Γ. Since the values of synthesized attributes in the forms *F*1*, . . . , Fk* are variables (by Definition [3.2](#_bookmark5)) and since these variables are unaffected by substitution *σin* the synthesized attribute in the resulting forms *Fjσin* are variables. The substitutions *σin* and *σout* substitute terms to the variables *x*1*, . . . , xk* appearing to the patterns and to the variables *y*1*, . . . , ym* respectively. Since *xi* appears in an input position in *P* , it can appear only in an output position in the forms !(*F*1)*, . . .*!(*Fk* ) and thus any variable of the term *σin* (*xi*) will appear in an output position in !(*Fiσin* ). Similarly, since *yi* appears in an input position in the form !(*s*(*u*1*, . . . , un*)(*y*1*, . . . , ym*)), it can only appear in an output position in !(*F* ) for the others forms *F* of Γ. Consequently any variable of the term *σout* (*yi*) will appear in an output position in !(*Fσout* ) for any equation *X* = *F* in Γ with *X* ƒ= *X*0. It follows that the application of a production cannot produce new occurrences of a variable in an input position and thus there cannot exist two occurrences *x*? of a same variable *x* in Γ′. ✷ *Prop*. [4.5](#_bookmark8)

✷

Thus applying an enabled production defines a binary relation on configurations.

**Definition 4.6.** *A configuration* Γ′ *is* ***directly accessible*** *from* Γ*, denoted by* Γ[ )Γ′*, whenever* Γ[*P/X*)Γ′ *for some production P enabled in node X of configuration* Γ*. Furthermore, a configu- ration* Γ′ *is* ***accessible*** *from configuration* Γ *when* Γ[∗)Γ′ *where* [∗) *is the reflexive and transitive closure of relation* [ )*.* ✷ *Def.* [*4.6*](#_bookmark8)

Inria

*A Grammatical Approach to Data-centric Case Management* 13

Recall that a substitution *σ* unifies a set of equations *E* if *tσ* = *t*′*σ* for every equations *t* = *t*′ in *E*. A substitution *σ* is more general than a substitution *σ*′ if *σ*′ = *σσ*′′ for some substitution *σ*′′. If a system of equations has a some unifier, then it has (up to an bijective renaming of the variables in *σ*) a *most general unifier*. In particular a set of equations of the

form {*xi* = *ti* | 1 ≤ *i* ≤ *n* } has a unifier if and only if it is acyclic. In this case, the corresponding solved form is its most general unifier.

**Proposition 4.7.** *If the left-hand side F* = *s*(*p*1*, . . . , pn*)(*u*1*, . . . , um*) *of a production P matches with the definition X* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) *of an open node X then the substitution σ* = match(*F, X*) *is the most general unifier of the set of equations* {*pi* = *di* | 1 ≤ *i* ≤ *n* } ∪

{*yj* = *uj* | 1 ≤ *j* ≤ *m* }*.*

In order to prove Proposition [4.7](#_bookmark9) we first recall some fact about unification.

**Recall 4.8** (on Unification)**.** *We consider sets E* = *E*? ⊎ *E*= *containing equations of two kinds. An equation in E , denoted as t* =? *u, represents a unification goal whose solution is a substitution σ such that tσ* = *uσ, i.e., substitution σ unifies terms t and u. E*= *contains only equations of the form x* = *t where variable x occurs only there, i.e., we do not have two equations with the same*

?

hal-00990007, version 1 - 13 May 2014

*variable in their left-hand side and such a variable cannot either occur in any right-hand side of an equation in E*=*. A solution to E is any substitution σ whose domain is the set of variables occurring in the right-hand sides of equations in E*= *such that the compound substitution made*

*of σ and the set of equations* {*x* = *tσ* | *x* = *t* ∈ *E* } *unifies terms t and u for any equation t* =? *u in E*?*. Two systems of equations are said to be equivalent when they have the same solutions. A unification problem is a set of such equations with E*= = ∅*, i.e., it is a set of unification goals. On the contrary E is said to be in solved form if E*? = ∅*, thus E defines a substitution which, by definition, is the most general solution to E. Solving a unification problem E consists in finding an equivalent system of equations E*′ *in solved form. In that case E*′ *is a most general unifier*

=

*for E.*

*Martelli and Montanari Unification algorithm [*[*18*](#_bookmark22)*] proceeds as follows. We pick up non de- terministically one equation in E*? *and depending on its shape apply the corresponding transfor- mation:*

? ? ?

1. *c*(*t*1*, . . . , tn*) = *c*(*u*1*, . . . , un*)*: replace it by equations t*1 = *u*1*, . . . , t*1 = *u*1*.*
2. *c*(*t , . . . , t* ) =? *c*′(*u , . . . , u* ) *with c* ƒ= *c*′*: halt with failure.*

1 *n* 1 *m*

1. *x* =? *x: delete this equation.*
2. *t* =? *x where t is not a variable: replace this equation by x* =? *t.*

*5. x* =?

*t where x* ƒ∈ var(*t*)*: replace this equation by x* = *t and substitute x by t in all other*

*equations of E.*

*6. x* =? *t where x* ∈ var(*t*) *and x* ƒ= *t: halt with failure.*

*The condition in (5) is the occur check. Thus the computation fails either if the two terms of an equation cannot be unified because their main constructors are different or because a potential solution of an equation is necessarily an infinite tree due to a recursive statement detected by the*

*occur check. System E*′ *obtained from E by applying one of these rules, denoted as E* ⇒ *E*′*, is clearly equivalent to E. We iterate this transformation as long as we do not encounter a*

*failure and some equation remains in E*?*. It can be proved that all these computations terminate*

RR n° 8528

14 *Badouel & al.*

*and either the original unification problem E has a solution (a unifier) and every computation terminates (and henceforth produces a solved set equivalent to E describing a most general unifier of E) or E has no unifier and every computation fails. We let*

*σ* = mgu({*ti* = *ui* | 1 ≤ *i* ≤ *n* }) iff ,*ti* = *ui*

?

| 1 ≤ *i* ≤ *n* , ⇒∗ *σ*

✷ *Recall* [*4.8*](#_bookmark9)

Note that (5) and (6) are the only rules that can be applied to solve a unification problem

of the form ,*yi* = *ui*

?

| 1 ≤ *i* ≤ *n* ,, where the *yi*

are distinct variables. The most general unifier

exists when the occur check always holds, i.e., rule (5) always applies. The computation amounts to iteratively replacing an occurrence of a variable *yi* in one of the right-hand side term *uj* by its definition *ui* until no variable *yi* occurs in some *uj* . This process terminates precisely when the relation *yi* ≻ *yj* ⇔ *yj* ∈ *ui* is acyclic. When this condition is met we say that the set of equations {*yi* = *ui* | 1 ≤ *i* ≤ *n* } *is acyclic* and we say that it *defines* the substitution *σ* = mgu({*yi* = *ui* | 1 ≤ *i* ≤ *n* }).

hal-00990007, version 1 - 13 May 2014

*Proof of Proposition* [*4.7*](#_bookmark9)*.*

If a production *P* of left-hand side *s*(*p*1*, . . . , pn*)(*u*1*, . . . um*) is triggered in node *X*0 defined by

*X*0 = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) then by Definition [4.4](#_bookmark8)

,*pi* = *d* | 1 ≤ *i* ≤ *n* , ∪ ,*y* = *u* | 1 ≤ *j* ≤ *m* , ⇒∗ *σ* ∪ ,*y* = *u σ* | 1 ≤ *j* ≤ *m* ,

? ? ?

*i j j in j j in*

using only the rules (1) and (5). Now

*σin* ∪ ,*yj* = *u σ*

?

| 1 ≤ *j* ≤ *m* , ⇒∗ *σ*

∪ mgu {*y*

= *u σ*

| 1 ≤ *j* ≤ *m* }

*j in*

*in j*

*j in*

by applying iteratively rule (5) if the set of equations {*yj* = *ujσin* | 1 ≤ *j* ≤ *m* } satisfies the occur check. Then *σin* + *σout* ⇒∗ *σ* again by using rule (5). ✷ *Prop*. [4.7](#_bookmark9)

✷

Note that the converse does not hold. Namely, one shall not deduce from Proposition [4.7](#_bookmark9) that the relation Γ[*P/X*0)Γ′ is defined whenever the left-hand side lhs(*P* ) of *P* can be unified with the definition def(*X*0*,* Γ) of *X*0 in Γ with

Γ′ = {*X*0 = *P* (*X*1*, . . . , Xk*)} where *X*1*, . . . , Xk* are nodes added to Γ′

∪ {*X*1 = *F*1*σ, . . . , Xk* = *Fkσ*}

∪ {*X* = *Fσ* | (*X* = *F* ) ∈ Γ ∧ *X* ƒ= *X*0 }

where *σ* = mgu(lhs(*P* )*,* def(*X*0*,* Γ)) is the corresponding most general unifier. Indeed, when unifying *s*(*d*1*, . . . , dn, y*1*, . . . , ym*) with *s*(*p*1*, . . . , pn, u*1*, . . . , um*) one may generate an equation of the form *x* = *t* where *x* is a variable in an inherited data *di* and *t* is an instance of a corresponding subterm in the associated pattern *pi*. This would correspond to a situation where information is sent to the context of a node through one of its inherited attribute! Otherwise stated some parts of the pattern *pi* are actually used to filtered out the incoming data value *di* while some other parts of the same pattern are used to transfert synthesized information to the context.

As already mentioned, an artifact is refined by applying a production at one of its open node. However we also need means to initiate cases. To this extent, we define interfaces for GAGs, that describe how services can initialize new artifacts.

Inria

*A Grammatical Approach to Data-centric Case Management* 15

**Definition 4.9.** *The* ***interface*** *of a guarded attribute grammar is given by a subset* I *of forms F* = *s*(*t*1*, . . . , tn*)(*x*1*, . . . , xm*)*, called its* ***services*** *where the synthesized positions are (distinct) variables x*1*, . . . , xm. The invocation of the service produces a new artifact reduced to a single*

*open node defined by F, it is associated with* ***initial configuration***

Γ0 = {*X*0 = *s*(*t*1*, . . . , tn*)(*x*1*, . . . , xm*)}

*An* ***accessible configuration*** *of a guarded attribute grammar is a configuration accessible from one of its initial configurations.* ✷ *Def.* [*4.9*](#_bookmark10)

**Example 4.10.** *The attribute grammar for the flattening of a binary tree (Example* [*3.1*](#_bookmark4)*) can be presented as a guarded attribute grammar with the following productions:*

Fork : *bin* (*x*)(*y*)←*bin* (*z*)(*y*) *bin*(*x*)(*z*)

Leaf*a*: *bin* (*x*)(Cons*a*(*x*))←

*together with service* Init(*x*) = *bin* (*Nil*)(*x*) *whose invocation creates a new binary tree and expects for the list of its leaves. In comparison with Example* [*3.1*](#_bookmark4) *we avoid the construction of the artificial node Root whose sole purpose was to initialize the inherited attribute of the tree at its root.*

hal-00990007, version 1 - 13 May 2014

✷ *Exple* [*4.10*](#_bookmark11)

# 5 Some Examples

In this section we illustrate the behaviour of guarded attribute grammars with three examples. Example [5.1](#_bookmark11) describes an execution of the attribute grammar of Example [3.1](#_bookmark4). The specifica- tion in Example [3.1](#_bookmark4) is actually an ordinary attribute grammar because the inherited attributes in the left-hand sides of productions are plain variables. This example shows how data are lazily produced and send in push mode through attributes. It also illustrates the role of the data links

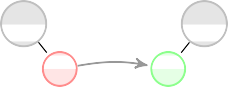
and their dynamic evolutions.

Example [5.2](#_bookmark14) illustrates the role of the guards by describing two processes acting as corou- tines. The first process sends forth a list of values to the second process and it waits for an acknowledgement for each message before sending the next one.

Example [5.3](#_bookmark15) justifies the role of the occur check.

**Example 5.1** (Example [3.1](#_bookmark4) continued)**.** *Let us consider the attribute grammar of Example* [*3.1*](#_bookmark4) *and the initial configuration* Γ0 = {*X*0 = *root*()(*x*)*, Y*0 = *toor* (*x*)()} *shown next*

*X*0 :: *root Y*0 :: *toor*



?

*x*

?

?

*x*

*The annotated version* !Γ0 = {!*F* | *F* ∈ Γ0 } *of configuration* Γ0 *is*

!Γ0 = .*X*0 = *root*()(*x*?)*, Y*0 = *toor* (*x*!)().

*The data link from x*? *to x*! *says that the list of the leaves of the tree —that will stem from node*

*X*0*— to be synthesized at node X*0 *should be forwarded to the inherited attribute of Y*0*.*

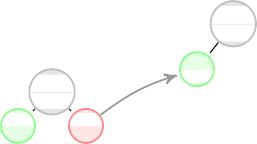
*This tree is not defined in the intial configuration* Γ0*. One can start developping it by ap- plying production* Root : *root*()(*u*) ← *bin* (*Nil*)(*u*) *at node X*0 :: *root. Actually the left-hand side root*()(*u*) *of production* Root *matches with the definition root*()(*x*) *of X*0 *with σin* = ∅ *and σout* = {*x* = *u*}*. Hence* Γ0[Root*/X*0)Γ1 *where the annotated configuration* !Γ1 *is given in Fig- ure* [*1*](#_bookmark12)*. Note that substitution σ*out = {*x* = *u*} *replaces the data link* (*x*?*, x*!) *by a new link* (*u*?*, u*!)

*with the same target and whose source has been moved from the synthesized attribute of X*0 *to the synthesized attribute of X*1*:*

RR n° 8528

16 *Badouel & al.*

*X*0 :: *root*



*X*1 :: *bin*

Root

?

? *Y*0 :: *toor*

*u*

*X*0 = Root(*X*1)

*X*1 = *bin*(Nil)(*u*?)

*Y*0 = *toor* (*u*!)( )

Nil ? *u*

Figure 1: Configuration Γ1

*u x u*

? *u x* [*σ*out)

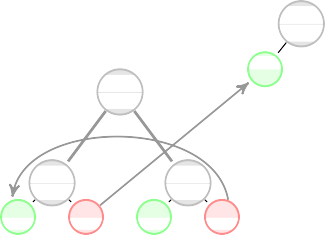


?

*u*

*The tree may be further refined by applying production* Fork : *bin* (*x*)(*y*) ← *bin* (*z*)(*y*) *bin*(*x*)(*z*) *at node X*1 :: *bin since its left-hand side bin* (*x*)(*y*) *matches with the definition bin* (Nil)(*u*) *of X*1 *with σin* = {*x* = *Nil*} *and σout* = {*u* = *y*}*. Hence* Γ1[Fork*/X*1)Γ2 *where* !Γ2 *is given in Figure* [*2*](#_bookmark12)*.*

*X*0 :: *root*



hal-00990007, version 1 - 13 May 2014

Root

? *Y*0 :: *toor*

*X*0 = Root(*X*1)

*y*

Fork

*X*1 = Fork(*X*11

*, X*12)

? *y* ? *z*

*z* ? Nil ?

*X*11 = *bin* (*z*!)(*y*?)

*X*12 = *bin* (*Nil*)(*z*?)

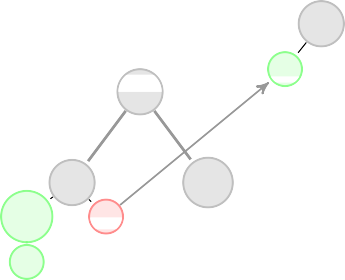
*Y*0 = *toor* (*y*!)( )

Figure 2: Configuration Γ2

*Production* Leaf*c* : *bin* (*x*)(Cons*c*(*x*)) ← *applies at node X*12 *since its left-hand side bin*(*x*)(Cons*c*(*x*))

*matches with the definition bin*(*Nil*)(*z*) *of X*12 *with σin* = {*x* = *Nil*} *and σout* = {*z* = Cons*c*(Nil)}*. Hence* Γ2[Leaf*c/X*12)Γ3 *where the annotated configuration* !Γ3 *is given in Figure* [*3*](#_bookmark12)*. As a result of*

*Y*0 :: *toor*



*X*0 :: *root*

Root

?

*y*

Fork

?

Leaf *c*

Cons*c*

?

*y*

Nil

*X*0 = Root(*X*1)

*X*1 = Fork(*X*11*, X*12)

*X*11 = *bin*(Cons*c*(Nil))(*y*?)

*X*12 = Leaf*c*

*Y*0 = *toor* (*y*!)( )

Figure 3: Configuration Γ3

*substitution σout* = {*z* = Cons*c*(Nil)} *the value* Cons*c*(Nil) *is transmitted through the link* (*z*?*, z*!)

*and this link disappears.*

*Production* Fork : *bin* (*x*)(*u*) ← *bin* (*z*)(*u*) *bin* (*x*)(*z*) *may apply at node X*11 *since its left-hand side bin* (*x*)(*u*) *matches with the definition bin* (Cons*c*(Nil))(*y*) *of X*11 *with σin* = {*x* = Cons*c*(Nil))} *and σout* = {*y* = *u*}*. Hence* Γ3[Fork*/X*1)Γ4 *with configuration* ?Γ4 *given in Figure* [*4*](#_bookmark13)*.*

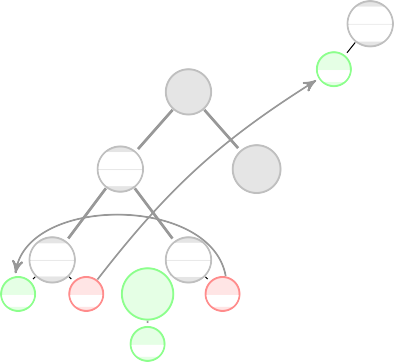
*Production* Leaf*a* : *bin*(*x*)(Cons*a*(*x*)) ← *applies at node X*111 *since its left-hand side bin* (*x*)(Cons*a*(*x*))

*matches with the definition bin* (*z*)(*u*) *of X*111 *with σin* = {*x* = *z*} *and σout* = {*u* = Cons*a*(*z*)}*. Hence* Γ4[Leaf*a/X*111)Γ5 *with configuration* !Γ5 *given in Figure* [*5*](#_bookmark13)*. Using substitution σout* =

Inria

*A Grammatical Approach to Data-centric Case Management* 17

*Y*0 :: *toor*



*X*0 :: *root*

Root

?

*u*

Fork

Fork

Leaf *c*

?

?

*z*

? Cons*c*

*u*

Nil

?

*z*

*X*0 = Root(*X*1)

*X*1 = Fork(*X*11*, X*12) *X*11 = Fork(*X*111*, X*112)

*X*111

= *bin* (*z*!)(*u*?)

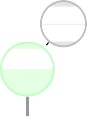
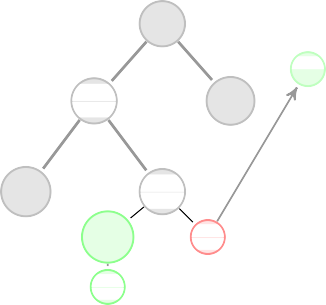
*X*112 = *bin* (Cons*c*(Nil))(*z*?)

*X*12 = Leaf*c*

*Y*0 = *toor* (*u*!)( )

Figure 4: Configuration Γ4

*Y*0 :: *toor*



*X*0 :: *root*

Root

?

Cons*a*

Fork

*z*

Fork

Leaf *c*

Leaf *a*

?

*z*

Cons*c*

?

Nil

*X*0 = Root(*X*1)

hal-00990007, version 1 - 13 May 2014

*X*1 = Fork(*X*11*, X*12) *X*11 = Fork(*X*111*, X*112)

*X*111 = Leaf*a*

*X*112 = *bin* (Cons*c*(Nil))(*z*?)

*X*12 = Leaf*c*

*Y*0 = *toor* (Cons*a*(*z*!))( )

Figure 5: Configuration Γ5

{*u* = Cons*a*(*z*)} *the data* Cons*a*(*z*) *is transmitted through the link* (*u*?*, u*!) *which, as a result, dis- appears. A new link* (*z*?*, z*!) *is created so that the rest of the list, to be synthesized in node X*112

*can later be forwarded to the inherited attribute of Y*0*.*

*Finally one can apply production* Leaf*b* : *bin* (*x*)(Cons*a*(*x*)) ← *at node X*112 *since its left-hand side matches with the definition bin* (Cons*c*(Nil))(*z*) *of X*112 *with σin* = {*x* = Cons*c*(Nil)} *and σout* = {*z* = Cons*b* (Cons*c*(Nil))}*. Therefore* Γ5[Leaf*b/X*112)Γ6 *with configuration* !Γ6 *given in Figure* [*6*](#_bookmark14)*. Now the tree rooted at node X*0 *is closed —and thus it no longer holds attributes—*

*and the list of its leaves has been entirely forwarded to the inherited attribute of node Y*0*. Note that the recipient node Y*0 *could have been refined in parallel with the changes of configurations just described.* ✷ *Exple* [*5.1*](#_bookmark11)

The above example shows that data links are used to transmit data in push mode from a source vertex *v* (the input occurrence *x*? of a variable *x*) to some target vertex *v*′ (an output occurrence *x*! of the same variable). These links (*x*! *, x*?) are *transient* in the sense that they disappear as soon as variable *x* gets defined by the substitution *σout* induced by the application of a production in some open node of the current configuration. If *σout* (*x*) is a term *t*, not reduced to a variable, with variables *x*1*, . . . , xk* then vertex *v*′ is refined by the term *t*[*x*! */xi*]

*i*

and new vertices *v*′ —associated with these new occurrences of *xi* in an output position— are

*i*

created. The original data link (*x*?*, x*!) is replaced by all the corresponding instances of (*x*?*, x*! ).

*i i*

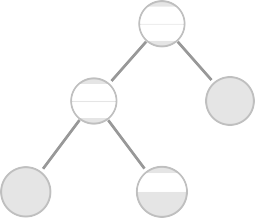
Consequently, a target is replaced by new targets which are the recipients for the subsequent

pieces of information (maybe none because no new links are created when *t* contains no variable).

RR n° 8528

18 *Badouel & al.*

*X*0 :: *root*



Root

Fork

Fork

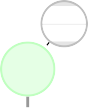
Leaf *c*

Leaf *a*

Leaf *b*

*Y*0 :: *toor*

*X*0 = Root(*X*1)



?

Cons*a*

Cons*b* Cons*c*

Nil

*X*1 = Fork(*X*11*, X*12) *X*11 = Fork(*X*111*, X*112)

*X*111 = Leaf*a X*112 = Leaf*b X*12 = Leaf*c*

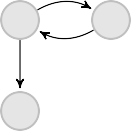
*Y*0 = *toor* (Cons*a*(Cons*b*(Cons*c*(Nil))))( )

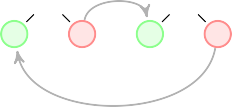
Figure 6: Configuration Γ6

If the term *t* is a variable *y* then the link (*x*?*, x*!) is replaced by the link (*y*?*, y*!) with the same target and whose source, the (unique) occurrence *x*? of variable *x*, is replaced by the (unique) occurrence *y*? of variable *y*. Therefore the direction of the flow of information is in both cases preserved: Channels can be viewed as “generalized streams” (that can fork or vanish) through which information is pushed incrementally.

hal-00990007, version 1 - 13 May 2014

**Example 5.2.** *Figure* [*7*](#_bookmark14) *shows a guarded attribute grammar that represents two coroutines com- municating through lazy streams. Each process alternatively sends and receives data. More precisely the second process send an acknowlegment (a b message) upon reception of a message send by the left process. Initially or after reception of an acknowlegment of its previous message the left process can either send a new message or terminate the communication.*

!*a*



"

*q*1

*q*′

2

? ?

*x* ? *y* ?

*y*

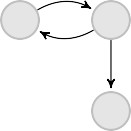
*q*1 *q*2

!stop ?*b*

*x*

!*b*

stop



*q*′

1

*q*′

2

?*a*

?

−

−

!*a* : *q*1 (*x*)(*a*(*y*)) ← *q*2 (*x*)(*y*)

!*b* : *q*′ ′

2 (*y*)(*x*)

?*b* : *q* (*b*(*x*))(*y*) ← *q* (*x*)(*y*)

1 (*y*)(*b*(*x*)) ← *q*2 (*y*)(*x*)

?*a* : *q*′ (*a*(*y*))(*x*) ← *q*′ (*y*)(*x*)

" : *q*0 ← *q*1 (*x*)(*y*) *q*′

2 1

!stop : *q*1 (*x*)(stop) ←

2 1

?stop : *q*′ (stop)(*y*) ←

2

Figure 7: Coroutines with lazy streams

*Production* !*a* : *q*1(*x*′)(*a*(*y*′)) ← *q*2(*x*′)(*y*′) *applies at node X*1 *of the configuration*

Γ1 = .*X* = *X*1 "*X*2 *, X*1 = *q*1 (*x*)(*y*)*, X*2 = *q*′

2 (*y*)(*x*).

*shown on the left of Figure* [*8*](#_bookmark15) *because its left-hand side q*1(*x*′)(*a*(*y*′)) *matches with the definition*

*q*1(*x*)(*y*) *of X*1 *with σin* = {*x*′ = *x*} *and σout* = {*y* = *a*(*y*′)}*. One obtains the configuration*

2(*a*(*y* ))(*x*)*, X*

= *q* (*x*)(*y*′).

*shown on the middle of Figure* [*8*](#_bookmark15)*. Production* ?*a* : *q*′ (*a*(*y*))(*x*′) ← *q*′ (*y*)(*x*′) *applies at node X*2

2 1

*of* Γ2 *because its left-hand side q*′ (*a*(*y*))(*x*′) *matches with the definition q*′ (*a*(*y*′))(*x*) *of X*2 *with*

2 2

*σin* = {*y* = *y*′} *and σout* = {*x* = *x*′}*. One obtains the configuration*

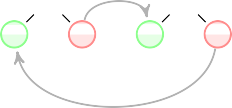
Γ3 = .*X* = *X*1 "*X*2 *, X*1 =!*a*(*X*11 )*, X*2 =?*a*(*X*21 )*, X*11 = *q*2 (*x*′)(*y*′)*, X*21 = *q*′ ′ ′

1 (*y* )(*x* ).

Inria

*A Grammatical Approach to Data-centric Case Management* 19

*x*



"

*q*1

*q*′

2

*x*

?

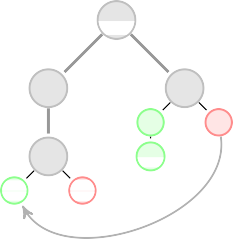
*y*

?

?

*y*

?



"

*q*′

2

!*a*

?

*q*2

*a*

*y*′

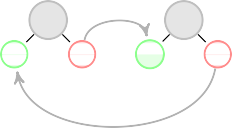
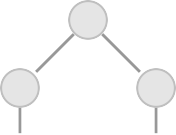
?

*x*

?

*y*′

?



"

!*a*

?*a*

*q*1

*q*′

1

?

*x*′ *y*′

?

?

*y*′

?

*x x*′

Figure 8: Γ1[!*a/X*1)Γ2[?*a/X*2)Γ3

*shown on the right of Figure* [*8*](#_bookmark15)*. The corresponding acknowlegment may then be send and received leading to configuration*

Γ5 = Γ ∪ .*X*111 = *q*1 (*x*)(*y*)*, X*211 = *q*′

2 (*y*)(*x*).

hal-00990007, version 1 - 13 May 2014

*where* Γ = {*X* = *X*1 "*X*2 *, X*1 =!*a*(*X*11 )*, X*2 =?*a*(*X*21 )*, X*21 =!*b*(*X*211 )*, X*11 =?*b*(*X*111 )}*. The process on*

*the left may decide to end the communication by applying production* !stop : *q*1(*x*′)(stop) ← *at*

*X*111 *with σin* = {*x*′ = *x*} *and σout* = {*y* = stop} *leading to configuration*

Γ6 = Γ ∪ .*X*111 =!stop*, X*211 = *q*′

2 (stop)(*x*).

*The reception of this message by the process on the right corresponds to applying production*

?stop : *q*′ (stop)(*y*) ← *at X*211 *with σin* = ∅ *and σout* = {*x* = *y*} *leading to configuration*

2

Γ7 = Γ ∪ {*X*111 =!stop*, X*211 =?stop}

*Note that variable x appears in an input position in* Γ6 *and has no corresponding output occurrence. This means that the value of x is not used in the configuration. When production* ?stop *is fired in node X*211 *variable y is substituted to x. Variable y has an output occurrence in production* ?stop *and no input occurence meaning that the corresponding output attribute is not defined by the semantic rules. As a consequence this variable simply disappears in the resulting configuration* Γ7 *. If variable x was used in* Γ6 *then the output occurrences of x would have been replaced by (output occurrences) of variable y which will remain undefined (no value will be substituted to y in subsequent transformations) until these*

*occurrences of variables may possibly disappear.* ✷ *Exple* [*5.2*](#_bookmark14)

We say that a production *P* is **triggered** in node *X* if substitution *σin* is defined, i.e., the patterns *pi* match the data *di*. As shown by the following example one can usually suspect an error in the specification when a triggered transition is not enabled due to the fact that the system of equations {*yj* = *ujσin* | 1 ≤ *j* ≤ *m* } is cyclic.

**Example 5.3.** *Let us consider the guarded attribute grammar given by the following productions:*

*P* : *s*0( )( ) ← *s*1(*a*(*x*))(*x*) *s*2(*x*)( )

*Q* : *s*1(*y*)(*a*(*y*)) ←

*R* : *s*2(*a*(*z*))( ) ←

*Applying production P*0 *in node X*0 *of configuration* Γ0 = {*X*0 = *s*0( )( )} *leads to configuration*

Γ1 = {*X*0 = *P* (*X*1*, X*2); *X*1 = *s*1(*a*(*x*))(*x*); *X*2 = *s*2(*x*)( )}

*Production Q is triggered in node X*1 *with σin* = {*y* = *a*(*x*)} *but the occur check fails because variable x occurs in a*(*y*)*σin* = *a*(*a*(*x*))*. Alternatively, we could drop the occur check and instead*

RR n° 8528

20 *Badouel & al.*

*adapt the fixed point semantics for attribute evaluation defined in [*[*3*](#_bookmark21)*,* [*19*](#_bookmark22)*] in order to cope with*

*infinite data structures. More precisely we could let σout be defined as the least solution of the system of equations* {*yi* = *ujσin* | 1 ≤ *j* ≤ *m* } *—assuming these equations are guarded, i.e., that there is no cycle of copy rules in the link graph of any accessible configuration—. In that case the infinite tree aω is substituted to variable x and the unique maximal computation associated with the grammar is given by the infinite tree P* (*Q, Rω* )*. In Definition* [*4.4*](#_bookmark8) *we have chosen to*

*restrict ourself to finite data structures which seems a reasonable assumption in view of the nature of systems we want to model. The occur check is used to avoid recursive definitions of attribute values. The given example, whose most natural interpretation is given by fixed point computation,*

*should in that respect be considered as ill-formed. And indeed this guarded attribute grammar is not sound (a notion presented in Section* [*7*](#_bookmark20)*) because the configuration* Γ *is not closed (it still contains open nodes), hence it represents a case that is not terminated. However it is a terminal configuration since it enables no production.* ✷ *Exple* [*5.3*](#_bookmark15)

# Distribution of a Guarded Attribute Grammar

The fact that triggered productions are not enabled can also impact the distributability of a grammar as shown by the following example.

hal-00990007, version 1 - 13 May 2014

**Example 6.1.** *Let us consider the GAG with the following productions:*

*P* : *s*( )( ) ← *s*1(*x*)(*y*) *s*2(*y*)(*x*) *Q* : *s*1(*z*)(*a*(*z*)) ←

*R* : *s*2(*u*)(*a*(*u*)) ←

*Production P is enabled in the configuration* Γ0 = {*X*0 = *s*( )( )} *with* Γ0[*P/X*0)Γ1 *where* Γ1 =

{*X*0 = *P* (*X*1*, X*2); *X*1 = *s*1(*x*)(*y*)*, X*2 = *s*2(*y*)(*x*)}*. In configuration* Γ1 *productions Q and R are enabled in nodes X*1 *and X*2 *respectively with*

Γ1[*Q/X*1)Γ2 where Γ2 = {*X*0 = *P* (*X*1*, X*2); *X*1 = *Q, X*2 = *s*2(*a*(*x*))(*x*)}

Γ1[*R/X*2)Γ3 where Γ3 = {*X*0 = *P* (*X*1*, X*2); *X*1 = *s*2(*a*(*y*))(*y*)*, X*2 = *R*}

*Now production R is triggered but not enabled in node X*2 *configuration* Γ2 *because of the cyclicity of* {*x* = *a*(*a*(*x*))}*. Similarly, production Q is triggered but not enabled in node X*3 *configuration* Γ3*. There is a conflict between the application of productions R and Q in configuration* Γ1*, which*

*makes this specification non-implementable in case nodes X*1 *and X*2 *have distinct locations.*

✷ *Exple* [*6.1*](#_bookmark16)

* 1. **Input Enabled Guarded Attribute Grammars**

Substitution *σin* , given by pattern matching, is monotonous w.r.t. incoming information and thus it causes no problem for a distributed implementation of a model. However substitution *σout* is not monotonous since it may become undefined when information coming from a distant location makes the match of output attributes a cyclic set of equations, as illustrated by example [6.1](#_bookmark16).

**Definition 6.2.** *A guarded attribute grammar is* ***input-enabled*** *if every production that is triggered in an accessible configuration is also enabled.* ✷ *Def.* [*6.2*](#_bookmark16)

We call the *substitution induced by a sequence* Γ[∗)Γ′ the corresponding composition of the various substitutions associated respectively with each of the individual steps in the sequence. If *X* is an open node in both Γ and Γ′, i.e., no productions are applied at *X* in the sequence, then we get *X* = *s*(*d*1*σ, . . . , dnσ*)(*y*1*, . . . , ym*) ∈ Γ′ where *X* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) ∈ Γ and *σ* is the substitution induced by the sequence.

Inria

*A Grammatical Approach to Data-centric Case Management* 21

**Proposition 6.3** (Monotony)**.** *Let* Γ *be an accessible configuration of an input-enabled GAG, X* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) ∈ Γ *and σ the substitution induced by some sequence starting from* Γ*. Then* Γ[*P/X*)Γ′ *implies* Γ*σ*[*P/X*)Γ′*σ.*

*Proof of Proposition* [*6.3*](#_bookmark16)*.*

Direct consequence of Definition [3.2](#_bookmark5) due to the fact that

1. match(*p, dσ*) = match(*p, d*)*σ*, and
2. mgu({*yj* = *ujσ* | 1 ≤ *j* ≤ *m* }) = mgu({*yj* = *uj* | 1 ≤ *j* ≤ *m* })*σ*.

The former is trivial and the latter follows by induction on the length of the computation of the most general unifier (relation ⇒∗ using rule (5) only). Note that the assumption that the guarded attribute grammar is input-enabled is crucial because in the general case it could happen that the set {*yj* = *ujσin* | 1 ≤ *j* ≤ *m* } satisfies the occur check whereas the set {*yj* = *uj* (*σin σ*) | 1 ≤ *j* ≤ *m* } does not satisfy the occur check. ✷ *Prop*. [6.3](#_bookmark16)

✷

Proposition [6.3](#_bookmark16) is instrumental for the distributed implementation of guarded attribute gram- mars. Namely it states that new information coming from a distant asynchronous location re- fining the value of some input occurrences of variables of an enabled production do not prevent from applying that production. Thus a production that is locally enabled can freely be applied regardless of information that might further refine the current local configuration. It means that conflict arises only from the existence of two distinct productions enabled in the same open node. Hence the only form of non-determinism corresponds to the decision of a stakeholder to apply one particular production among those enabled in a configuration. This is expressed by the following confluence property.

hal-00990007, version 1 - 13 May 2014

**Corollary 6.4.** *Let* Γ *be an accessible configuration of an input enabled GAG. If* Γ[*P/X*)Γ1 *and*

Γ[*Q/Y* )Γ2 *with X* ƒ= *Y then* Γ2[*P/X*)Γ3 *and* Γ1[*Q/Y* )Γ3 *for some configuration* Γ3*.*

Note that, by Corollary [6.4](#_bookmark17), the artifact contains a full history of the case in the sense that one can reconstruct from the artifact the complete sequence of applications of productions leading to the resolution of the case (up to the commutation of independent elements in the sequence). We might have considered a more symmetrical presentation in Definition [3.2](#_bookmark5) by allowing patterns for synthesized attributes in the right-hand sides of productions with the effect of creating forms in a configuration with patterns in their co-arguments. These patterns express constraints on synthesized values. This extension could be acceptable as long as one sticks to purely centralized models. However, as soon as one wants to distribute the model on an asynchronous architecture, one cannot avoid such a constraint to be further refined due to a transformation occurring in a distant location. Then the monotony property (Proposition [6.3](#_bookmark16)) is lost: a locally enabled production can later be disabled when a constraint on a synthesized value gets a refined value. This is why we required synthesized attributes in the right-hand side of a production to be given by plain variables in order to prohibit the expression of constraints

on synthesized values.

It is difficult to verify input-enabledness as the whole set of accessible configurations are involved in this condition. Nevertheless one can find a sufficient condition for input enabledness, similar to the strong non-circularity of attribute grammars [[4](#_bookmark21)], that can be checked by a simple fixed-point computation.

**Definition 6.5.** *Let s be a sort of a guarded attribute grammar with n inherited attributes and m synthesized attributes. We let* (*j, i*) ∈ *SI*(*s*) *where* 1 ≤ *i* ≤ *n and* 1 ≤ *j* ≤ *m if exists X* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) ∈ Γ *where* Γ *is an accessible configuration and yj* ∈ *di. If P is a*

RR n° 8528

22 *Badouel & al.*

*production with left-hand side s*(*p*1*, . . . , pn*)(*u*1*, . . . , um*) *we let* (*i, j*) ∈ *IS*(*P* ) *if exists a variable x* ∈ var(*P* ) *such that x* ∈ var(*di*) ∩ var(*uj* )*. The guarded attribute grammar tt is said to be* ***acyclic*** *if for every sort s and production P whose left-hand side is a form of sort s the graph tt*(*s, P* ) = *SI*(*s*) ∪ *IS*(*P* ) *is acyclic.* ✷ *Def.* [*6.5*](#_bookmark17)

**Proposition 6.6.** *An acyclic guarded attribute grammar is input-enabled.*

*Proof.* Suppose *P* is triggered in node *X* with substitution *σin* such that *yj* ∈ *uiσin* then (*i, j*) ∈ *tt*(*s, P* ). Then the fact that occur check fails for the set {*yj* | 1 ≤ *j* ≤ *m* } entails that one can find a cycle in *tt*(*s, P* ). ✷ *Prop*. [6.6](#_bookmark18)

✷

Relation *SI*(*s*) still takes into account the whole set of accessible configurations. The following definition provides an overapproximation of this relation given by a fixed point computation.

**Definition 6.7.** *The* ***graph of local dependencies*** *of a production P* : *F*0 ← *F*1 · · · *Fℓ is the directed graph ttLD*(*P* ) *that records the data dependencies between the occurrences of attributes*

*given by the semantics rules. We designate the occurrences of attributes of P as follows: we let k*(*i*) *(respectively k*(*j*)*) denote the occurrence of the ith inherited attribute (resp. the jth synthesized attribute) in Fk. If s is a sort with n inherited attributes and m synthesized attributes we define the relations IS*(*s*) *and SI*(*s*) *over* [1*, n*] × [1*, m*] *and* [1*, m*] × [1*, n*] *respectively as the least relations such that :*

hal-00990007, version 1 - 13 May 2014



1. *SI*(*s*) = *SI*(*s*) *if s is an axiom, i.e., it is given by the set of pairs* (*j, i*) *such that yj* ∈ var(*di*) *for some service F* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) *of sort s in the interface of the guarded attribute grammar.*
2. *For every production P* : *F*0 ← *F*1 · · · *Fℓ where form Fi is of sort si and for every k* ∈ [1*, ℓ*]

.(*j, i*) . (*k*(*j*)*, k*(*i*)) ∈ *ttLD*(*P* )*k* . ⊆ *SI*(*sk*)

.

*where graph ttLD*(*P* )*k is given as the transitive closure of*

*ttLD*(*P* ) ∪ ,(0(*j*)*,* 0(*i*)) . (*j, i*) ∈ *SI*(*s*0),

.

.

∪ ,(*k*′(*i*)*, k*′(*j*)) . *k*′ ∈ [1*, ℓ*]*, k*′ ƒ= *k,* (*i, j*) ∈ *IS*(*sk*′ ),

.

.

1. *For every production P* : *F*0 ← *F*1 · · · *Fℓ where form Fi is of sort si*

.(*i, j*) . (0(*i*)*,* 0(*j*)) ∈ *ttLD*(*P* )0 . ⊆ *IS*(*s*0)

.

*where graph ttLD*(*P* )0 *is given as the transitive closure of*

*ttLD*(*P* ) ∪ ,(*k*(*i*)*, k*(*j*)) . *k* ∈ [1*, ℓ*]*,* (*i, j*) ∈ *IS*(*sk*),

.

.

*The guarded attribute grammar tt is said to be* ***strongly-acyclic*** *if for every sort s and production*

*P whose left-hand side is a form of sort s the graph tt*(*s, P* ) = *SI*(*s*) ∪ *IS*(*P* ) *is acyclic.*

✷ *Def.* [*6.7*](#_bookmark18)

**Proposition 6.8.** *A strongly-acyclic guarded attribute grammar is acyclic and hence input- enabled.*

Inria

*A Grammatical Approach to Data-centric Case Management* 23

*Proof.* The proof is analog to the proof that a strongly non-circular attribute grammar is non- circular and it goes as follows. We let (*i, j*) ∈ *IS*(*s*) when var(*diσ*) ∩ var(*yjσ*) ƒ= ∅ for some form *F* = *s*(*d*1*, . . . , dn*)(*y*1*, . . . , ym*) of sort *s* and where *σ* is the substitution induced by a firing sequence starting from configuration {*X* = *F* }. Then we show by induction on the length of the firing sequence leading to the accessible configuration that *IS*(*s*) ⊆ *IS*(*s*) and *SI*(*s*) ⊆ *SI*(*s*).

✷ *Prop*. [6.8](#_bookmark18)

✷

Note that the following two inclusions are strict

strongly-acyclic GAGs Ç acyclic GAGs Ç input enabled GAGs Indeed the reader may easily check that the guarded attribute grammar

. *A*(*x*)(*z*) ← *B*(*a*(*x, y*))(*y, z*)

*B*(*a*(*x, y*))(*x, y*) ←

is cyclic and input-enabled whereas guarded attribute grammar with productions

hal-00990007, version 1 - 13 May 2014

 *A*(*x*)(*z*) ← *B*(*y, x*)(*z, y*)

 *A*(*x*)(*z*) ← *B*(*x, y*)(*y, z*)

 *B*(*x, y*)(*x, y*) ←

is acyclic but not strongly-acyclic. Attribute grammars arising from real situations are almost always strongly non-circular so that this assumption is not really restrictive. Similarly we are confident that most of the guarded attribute grammars that we shall use in practise will be input- enabled and that most of the input-enabled guarded attribute grammars are in fact strongly- acyclic. Thus most of the specifications are distributable and most of those can be proved so by checking the strong non-circularity condition.

## Distribution of an Input Enabled Guarded Attribute Grammar

The principle of a distribution of a GAG on a set of locations is as follows: Each location maintains a local configuration, and subscribes to results provided by other locations. Productions are applied locally. When variables are given a value by a production, the location that computed this value sends messages to the locations that subscribed to this value. Messages are simply equations defining the value of a particular variable. Upon reception of a messages, a subscriber updates its local configuration, and may in turn produce new messages.

More formally, a GAG can be distributed by specifying a partition *S* = ⊎1≤*ℓ*≤*pSℓ* of the set of sorts. The projections Γ*ℓ*, called the local configurations associated with sites *Sℓ*, are defined as follows. Each site *Sℓ* has a namespace *ns*(*Sℓ*) used for the nodes *X* whose sorts are in *Sℓ* and for the variables *x* representing attributes of these nodes but also for references to variables belonging to distant sites (subscriptions). Hence we have name generators that produce unique identifiers for each newly created variable for each site. For each equation *X* = *P* (*X*1*, . . . , Xn*) with *X* :: *s* and *Xi* :: *si* we insert equation *X* = *P* (*X*1*, . . . , Xn*) in Γ*ℓ* where *s* ∈ *Sℓ* and variable *Xi* is *Xi* if *si* ∈ *Sℓ* or is a new variable in the namespace of *Sℓ* if *si* ∈ *Sℓ*′ with *ℓ*′ ƒ= *ℓ*. In the latter case we add equation *Xi* = *Xi* in Γ*ℓ*. Similarly for each equation *X* = *s*(*t*1*, . . . , tn*)(*y*1*, . . . , ym*) in Γ we add equation *X* = *s*(*t*1*, . . . , tn*)(*y*1*, . . . , ym*) in Γ*ℓ* where *s* ∈ *Sℓ* and *t* is obtained by replacing each variable *x* in term *t* by *x* where variable *x* is *x* if *x* :: *s*′ with *s*′ ∈ *Sℓ* else is

a new variable in the namespace of *Sℓ*. In the latter case one adds equation *x* = *x*, called a

**subscription**, to Γ*ℓ*′ . Similarly for the variables *yj* . Hence a local configuration contains the

RR n° 8528

24 *Badouel & al.*

usual equations associated with their closed and open nodes (and containing only local variables) together with equations of the form *X* = *Y* and *y* = *x* where *x* and *X* are local names and *y* and *Y* belongs to distant sites. Clearly the global configuration can be recovered as Γ = Γ1 ⊕ · · · ⊕ Γ*n* where operator ⊕ consists in taking the union of the systems of equations given as arguments and simplifying the resulting system by elimination of the copy rules: we drop each equation of the form *X* = *Y* (respectively *y* = *x*) and replace each occurrence of *X* by *Y* (resp. of *y* by *x*). Therefore the global configuration Γ may be identified with the vectors of local configurations (Γ1*, . . . ,* Γ*p*).

Each production can then be locally applied: we write Γ *P*−*/*→*X* Γ′

*ℓ*

when application of pro-

*M ℓ*

duction *P* at node *X* results in a new configuration Γ′

*ℓ*

and the sending of a set of messages

*M* .

More formally, Γ *P*−*/*→*X* Γ′

*ℓ*

when *X* = *s*(*t*1*, . . . , tn*)(*y*1*, . . . , ym*) ∈ Γ*ℓ* and *P* = *F* ← *F*1 · · · *Fk* is

*M ℓ*

a production whose left-hand side matches with *X* and

*ℓ* = {*X* = *P* (*X*1*, . . . , Xk*)} where *X*1*, . . . , Xk* are new names in *ns*(*Sℓ*)

Γ′

∪ {*Xi* = *Fiσ* | *Xi* :: *si* and *si* ∈ *Sℓ* }

hal-00990007, version 1 - 13 May 2014

∪ {*X*′ = *Fσ* | (*X*′ = *F* ) ∈ Γ*ℓ* ∧ *X*′ ƒ= *X* }

∪ {*y*′ = *yjσ* | (*y*′ = *yj* ) ∈ Γ*ℓ* and *yjσ* is a variable}

*M* = {*Xi* = *Fiσ* | *Xi* :: *si* and *si* ƒ∈ *Sℓ* }

∪ {*y*′ = *yjσ* | (*y*′ = *yj* ) ∈ Γ*ℓ* and *yjσ* not a variable}

∪ *Mσ*

where *σ* = match(*F, X*), and *σ* is the relativization of *σ* to location *ℓ*, it generates a set of messages *Mσ* . This relation means that applying production *P* at *X* in site *Sℓ* generates messages *M* send to distant sites. The reception of a message may generate new messages and is described by relation Γ −*m*→ Γ′ where

*ℓ*

*M ℓ*

1. If *m* = {*X* = *s*(*t*1*, . . . , tn*)(*y*1*, . . . , yq* )} with *X* ∈ *ns*(*Sℓ*′ ), *s* ∈ *Sℓ* with *ℓ*′ ƒ= *ℓ* then Γ′ =

*ℓ*

Γ*ℓ* ∪ .*X* = *s*(*t*1*, . . . , tn*)(*y*1*, . . . , yq* ). ∪ {*yj* = *yj* | 1 ≤ *j* ≤ *q* } where *X*, the variables *x* for

*x* ∈ *var*(*ti*) and the variables *yj* are new names in *ns*(*Sℓ*) and *t* = *t*[*x/x*], and *M* =

{*x* = *x* | *x* ∈ *var*(*ti*) 1 ≤ *i* ≤ *n* } ∪ .*X* = *X*..

1. If *m* = {*x* = *t*} with *x* ∈ *ns*(*Sℓ*) then Γ′ = Γ*ℓ*[*x* = *t*[*y/y*]] where *y* are new names in *ns*(*Sℓ*)

*ℓ*

associated with the variables *y* in *t* and *M* = {*y* = *y* | *y* ∈ *var*(*t*)}.

1. If *m* = (*X* = *Y* ) with *X* ∈ *ns*(*Sℓ*) then Γ′ = Γ*ℓ* ∪ {*X* = *Y* } and *M* = ∅.

*ℓ*

1. If *m* = (*y* = *x*) with *x* ∈ *ns*(*Sℓ*) then Γ′ = Γ*ℓ* ∪ {*y* = *x*} and *M* = ∅.

*ℓ*

The global dynamics of the system can then be derived as follows, where *e* stands for *P/X*

or a message *m*:

*e e*

1. If Γ*ℓ* −→ Γ′ then Γ =⇒ Γ′ with Γ*ℓ*′ = Γ′ for *ℓ*′ ƒ= *ℓ*.

*M ℓ M ℓ*′

1. If Γ =*e*⇒ Γ′ and Γ′ =*m*⇒ Γ′′ for *m* ∈ *M* then Γ

=*e*⇒

Γ′′

*M M* ′ *M* \{*m*}∪*M* ′

Input-enabled GAGs possess useful properties with respect to distribution, namely messages consumptions and application of productions commute, as shown in the following proposition:

**Proposition 6.9.** *For an input-enabled guarded attribute grammar:*

1. *If* Γ *P*=*/*⇒*X* Γ′ *then there exists a substitution σ*

*M*

*M*

*such that* Γ *P*=*/*⇒*X* Γ′ σ *.*

∅

*M*

Inria

*A Grammatical Approach to Data-centric Case Management* 25

1. Γ[*P/X*)Γ′ *if and only if* Γ *P*=*/*⇒*X* Γ′

∅

1. *Let* Γ *P*=1 */*⇒*X*1 Γ1 *and* Γ *P*=2*/*⇒*X*2 Γ2 *with X*

1

ƒ= *X*2*. One can assume w.l.o.g that M*1

*and M*2

*have*

*M*1 *M*2

*no common variables (the name generator chooses different names for the new variables in*

*both cases). Then the diagram below , where* ~ *denotes messages consumption commutes.*

♠ Γ ◗◗

♠♠ ♠♠ 1

*P*1 */X*1

♠♠♠♠♠

◗◗◗◗◗*P*2*/X*2

Γ *σ* Γ

♠♠

♠ *M*

,t♠♠♠

◗◗◗◗◗◗

,¸ Γ Γ *σ*

*M*2 ◗◗◗

1 *M*1 ,r o/ o/

1 ,,

,,,*P*2 */X*2

,,

*P*1 */X*1

2

♥♥♥♥

♥

/o 2 *M*2

*M*2 *P*2 */X*2

,,,

♥♥♥

*M*1 *P*1 */X*1

,, *M*2

Γ′

,,,,,

♥♥♥♥*M*1

,t♥ ′

♥♥

1 ,, Γ3 Γ2

¸,

.›

,, ,t ¸, ¸, ¸,

¸, ¸, ¸, ¸,

Γ3(*σM*1 ∪ *σM*2 )

= Γ′ *σM* = Γ′ *σM*

hal-00990007, version 1 - 13 May 2014

2 1 1 2

Intuitively, proposition [6.9](#_bookmark19) and in particular (3) mean that distribution does not affect the global behavious of an input-enabled GAG.

*Proof of Proposition* [*6.9*](#_bookmark19)*.*

We first prove (1): whenever Γ *P*=*/*⇒*X* Γ′ then there exists a substitution *σ*

*M*

*M*

such that Γ *P*=*/*⇒*X* Γ′ σ .

∅

*M*

Let us assume that Γ *P*=*/*⇒*X* Γ′, and examine how consuming messages in *M* = {*m , . . . , m* } affects

1 *q*

*M*

Γ′. Messages can be of several kinds :

* if *mi* = (*y* = *x*) ( or *mi* = (*X* = *Y* ) ) then consuming the message results in adding an equation *σmi* = {*y* = *x*} (resp. *σmi* = {*X* = *Y* }) to the local configuration that receives this message, and generates no new message.
* if *mi* = {*x* = *t*}, then consumption of the message results in production of new variables, and a new (finite) set of messages *Mi* that are all of the form {*y*¯ = *y*} and can then be consumed without producing new messages by the location that has subscribed to this value. We can denote by *σi* the substitution that replaces every *x* in the local configuration that receives *mi*.
* if *mi* is of the form {*X* = *s*(*t*1*, . . . , tn*) *< y*1*, . . . , yn >*}, then consuming *mi* results in adding new equations to the local configuration that receives it, and generating a set of messages *Mi* = {*mi,*1*, mi,q* }, that are of the form {*y*¯ = *y*} and {*X* = *X*¯ } and can hence be consumed by the location that will receive them without generating new messages.

These observations show that, after application of a production, message consumption is a finite

process. We have Γ *P*=*/*⇒*X* Γ′=*m*⇒1 Γ =*M*⇒1 Γ′ *. . . m*|⇒*M* | Γ

1 =

*M*

=⇒ Γ′

|*M* |

*M M*1

1

∅ *M*|*M* |

|*M* |

∅

|*M* |

Now, the difference between each Γ*i* and Γ′ is a set of subscriptions, that are appended to some local configurations, and erased during the step. Therefore, the global configurations Γ*i*

*i*

and Γ′ are identical. Similarly, we have Γ1 = Γ′*σm* , and Γ*i* = Γ′

*σm* for every *i* ∈ [2*,* |*M* |].

*i* 1 *i*−1 *i*

′ ′

*P/X*

Hence, the substitution *σM* = *σm*1 *. . . σ*|*M*| is such that Γ *σM* = Γ|*M*|. Hence Γ =⇒ Γ′ σ*M* .

∅

We now give a proof for (2). We have to establish the following equivalence: Γ[*P/X*)Γ′

if and only if Γ *P*=*/*⇒*X* Γ′. First, whenever Γ *P*=*/*⇒*X* Γ′, then, by definition, Γ[*P/X*)Γ′. Conversely,

∅ ∅

RR n° 8528

26 *Badouel & al.*

*P/X P/X*

Γ[*P/X*)Γ′ implies the existence of *ℓ* and *M* such that Γ*ℓ* −→ Γ1. Thus, by definition Γ =⇒ Γ1 (with

Γ1 ′

*M ℓ M*

*P/X* 1

*ℓ*′ = Γ*ℓ*′ for *ℓ*

ƒ= *ℓ*). Hence, by (1) we have Γ =⇒ Γ σ*M* . Note that productions applications

∅

are deterministic, and messages consumption too. Hence, it suffices to prove Γ′ := Γ1*σM* to

obtain the desired result. In fact the nodes replacement performed to obtain Γ′ and Γ1 are identical, since the same production is applied at the same node. Let us denote by Γ[*P/X*] the configuration obtained by replacement of *X*. We have to show the equality Γ′ = Γ[*P/X*]*σ* = Γ[*P/X*]*σlσM* , where *σ* is the usual substitution applied during production application, *σl* is the substitution resulting from applying production locally to Γ*l*, and *σM* is the substitution obtained by consumption of messages in *M* . Now, one can notice that all substitutions in *σM* replace a variable *y* by a term *t* whenever *y* is a subscription to some value produced in Γ*l*. The effect is exactly the same as applying *σ* at nodes that differ from *X* in Γ. As additional subscription generated by messages consumption is not considered in the product, we have Γ′ = Γ[*P/X*]*σ* = Γ[*P/X*]*σlσM* .

The last statement, (3), expresses the commutativity of the following diagram:

♠ Γ ◗◗

♠♠ ♠♠ 1

hal-00990007, version 1 - 13 May 2014

*P*1*/X*1

♠♠♠♠♠

◗◗◗◗◗*P*2 */X*2

Γ *σ* Γ

♠♠

♠ *M*

,t♠♠♠

◗◗◗◗◗◗

,¸ Γ Γ *σ*

*M*2 ◗◗◗

1 *M*1 ,r o/ o/

1 ,,

,,,*P*2*/X*2

,,

*P*1 */X*1

2

♥♥♥♥

♥

/o 2 *M*2

*M*2 *P*2*/X*2

,,,

♥♥♥

*M*1 *P*1*/X*1

,, *M*2

Γ′

,,,,,

♥♥♥♥*M*1

,t♥ Γ′

♥♥

,,

1 ,, Γ3 2

¸,

.›

,, ,,

,t ¸, ¸, ¸,

¸, ¸, ¸, ¸,

Γ3(*σM*1 ∪ *σM*2 )

= Γ′ *σM* = Γ′ *σM*

2 1 1 2

We first consider the commutativity of the center:

Γ *P*=1 */*⇒*X*1 Γ1 *P*=2*/*⇒*X*2 Γ3 commutes into Γ *P*=2 */*⇒*X*2 Γ2 *P*=1*/*⇒*X*1 Γ3 .

*M*1 *M*2 *M*2 *M*1

This follows directly from the properties of input-enabled grammars: Γ3 is simply Γ where both open nodes *X*1 and *X*2 have been replaced respectively by the closed nodes *X*1 = *P*1(*Y* 1*, . . . , Y* 1 )

1 *m*

and *X*2 = *P*2(*Y* 2*, . . . , Y* 2) (since *M*1 and *M*2 have no common variables they are unaffected by

1 *k*

each other).

*P*1 */X*1

Let us consider the left hand-side of the diagram. From (1), we have that Γ

s, Γ1

*M*1

/o s, Γ1*σM*1 .

And by (2), this implies that Γ[*P*1*/X*1)Γ1*σM*1 . Using Proposition [6.3](#_bookmark16), and the fact that Γ1[*P*2*/X*2)Γ1,

1

whe have that *P*2 is triggered and enabled in Γ1*σM*1 . Hence, Γ1σ*M*1

*P*=2*/*⇒*X*2 *M*2

Γ′ . Using again Propo-

sition [6.3](#_bookmark16), configuration Γ′

1

is simply Γ3*σM*1 .

*P*2 */X*2 ′ ′ ′

Furthermore, from (1), we have that: Γ1*σM*1

*M*2

s, Γ1

/o s, Γ1*σM*2 , since Γ1 = Γ3*σM*1 , we

have: Γ′ *σM*

= Γ3*σM σM* . By a symmetric argument, we obtain: Γ′ *σM*

= Γ3*σM σM* . Since

1 2 1 2 2 1 2 1

these substitutions have disjoint support, we have *σM*1 *σM*2 = *σM*2 *σM*1 = *σM*1 ∪ *σM*2 . Thus, we have: Γ3*σM*1 *σM*2 = Γ3*σM*2 *σM*1 = Γ3*σM*1 ∪ *σM*2 .

Finally Γ3

/o s, Γ3(*σM*1 ∪ *σM*2 ) follows from the definitions of *σM*1 and *σM*2 .

✷ *Prop*. [6.9](#_bookmark19)

✷

Inria

*A Grammatical Approach to Data-centric Case Management* 27

# Soundness

A specification is sound if every case can reach completion no matter how its execution started.

**Definition 7.1.** *Let a guarded attribute grammar be given with its interface (Definition* [*4.9*](#_bookmark10)*). A* ***case*** *is an instanciation c* = *s*(*t*1*σ, . . . , tnσ*)(*x*1*, . . . , xm*) *of a service s*(*t*1*, . . . , tn*)(*x*1*, . . . , xm*) *where σ is a substitution such that* var(*ti*)*σ* ⊆ {*x*1*, . . . , xm*}*. Stated otherwise a case is, but for the variables with a synthesized value, a closed instance of a service. It means that it is a service call*

*which already contains all the information coming from the environment of the guarded attribute grammar. A configuration is* ***closed*** *if it contains only closed nodes. A guarded attribute grammar is* ***sound*** *if a closed configuration is accessible from any configuration* Γ *accessible from the initial configuration* Γ0(*c*) = {*X*0 = *c*} *associated with a case c.* ✷ *Exple* [*7.1*](#_bookmark20)

Let *γ* denote the set of configurations accessible from the initial configuration of some case. We consider the finite sequences (Γ*i*)0*<i*≤*n* and the infinite sequences (Γ*i*)0*<i<ω* of configurations in *γ* such that Γ*i*[ )Γ*i*+1. A finite and maximal sequence is said to be **terminal**, i.e., a terminal sequence leads to a configuration that enables no production. Soundness can the be rephrased by the two following conditions.

hal-00990007, version 1 - 13 May 2014

1. Every terminal sequence leads to a closed configuration.
2. Every configuration on an infinite sequence also belongs to some terminal sequence. Soundness can unfortunately be proved indecidable by a simple encoding of Minsky machines.

**Proposition 7.2.** *Soundness of guarded attribute grammar is undecidable.*

*Proof.* We consider the following presentation of the Minsky machines. We have two registers *r*1 and *r*2 holding integer values. Integers are encoded with the constant **zero** and the unary operator **succ**. The machine is given by a finite list of instructions *instri* for *i* = 1*, . . . , N* of one of the three following forms

1. **INC(r,i):** increment register *r* and go to instruction *i*.
2. **JZDEC(r,i,j):** if the value of register *r* is 0 then go to instruction *i* else decrement the value of the register and go to *j*.
3. **HALT:** terminate.

We associate such a machine with a guarded attribute grammar whose sorts corresponds bijec- tively to the lines of the program, (i.e., *S* = {*s*1*, . . . , sN* }) with the following encoding of the program instructions by productions:

1. If *instrk* = INC(*r*1*, i*) then add production

Inc(*k,* 1*, i*) : *sk*(*x, y*) ← *si*(succ(*x*)*, y*)

1. If *instrk* = INC(*r*2*, i*) then add production

Inc(*k,* 2*, i*) : *sk*(*x, y*) ← *si*(*x,* succ(*y*))

1. If *instrk* = JZDEC(*r*1*, i, j*) then add the productions

Jz(*k,* 1*, i*) : *sk*(zero*, y*) ← *si*(zero*, y*)

Dec(*k,* 1*, j*) : *sk*(succ(*x*)*, y*) ← *sj* (*x, y*)

RR n° 8528

28 *Badouel & al.*

1. If *instrk* = JZDEC(*r*2*, i, j*) then add the productions

Jz(*k,* 2*, i*) : *sk*(*x,* zero) ← *si*(*x,* zero)

Dec(*k,* 2*, j*) : *sk*(*x,* succ(*y*)) ← *sj* (*x, y*)

1. If *instrk* = HALT then add production

Halt(*k*) : *sk*(*x, y*) ←

Since there is a unique maximal firing sequence from the initial configuration Γ0 = {*X*0 = *s*1(zero*,* zero)} the corresponding guarded attribute grammar is sound if and only if the computation of the cor- responding Minsky machine terminates. ✷ *Prop*. [7.2](#_bookmark20)

✷

1. **Conclusion**

hal-00990007, version 1 - 13 May [2014](#_bookmark20)

To conclude we assess our model and highlight some research directions.

* 1. **Assessment of the model**

In a nutshell *the model of GAGs provides a modular, declarative, user-centric, data-driven, dis- tributed and reconfigurable model of case management*.

**Concurrency.** The lifecycle of a business artifact is implicitly represented by the grammar productions. A production decomposes a task into new subtasks and specifies constraints between their attributes in the form of the so-called semantic rules. The subtasks may then evolve independently as long as the semantic rules are satisfied. The order of execution, which may depend on value that are computed during process execution, need not (and cannot in general) be determined statically. For that reason, GAGs allow as much concurrency as needed. In comparison, models in which the lifecycle of artifacts are represented by finite automata constrain concurrency among tasks in an artificial way.

**Modularity.** The GAG approach also facilitates a modular description of business processes. For instance when a referee has accepted to produce a report, one need not care about the sub- process dedicated to the actual production of the report. In the example depicted on Table [3](#_bookmark3), making a review report was modeled by a single production, but one can imagine similar sit- uations where computing some synthesized information is given by a large set of rules used to collect and assemble information arising from various sources. However, following a top-down ap- proach, one simply introduces an attribute in which this report should eventually be synthesized and delegate the actual production of the expected outcome to an additional set of rules. The identification of the different roles involved in the business process also contributes to enhance modularity. Finally, some techniques borrowed from attribute grammars, like descriptional com- position [[10](#_bookmark21), [11](#_bookmark22)], decomposition by aspects [[30](#_bookmark23), [29](#_bookmark23)] or higher-order attribute grammars [[28](#_bookmark23)], may also contribute to better modular designs.

**Reconfiguration.** The workflow can be reconfigured at run time: New business rules (pro- ductions of the grammar) can be added to the system without disturbing the current cases. By contrast, run time reconfiguration of workflows modeled by Petri nets (or similar models) is a

Inria

*A Grammatical Approach to Data-centric Case Management* 29

complex issue [[20](#_bookmark22), [9](#_bookmark21)]. One can also add “macro productions” corresponding to specific compo- sitions of productions. For instance if the Editor-in-chief wants to handle the evaluation of a paper, he can decide to act as an associate editor and as a referee for this particular submission. However, this means forwarding the corresponding case to himself as an associate editor and then asking himself as a referee if he is willing to write a report. A more direct way to model this decision is to encapsulate these steps in a compound macro production that bypasses the intermediate communications. More generally compound rules can be introduced for handling unusual behaviors that deviates from the nominal workflow.

**Logged information.** When a case is terminated, the corresponding artifact collects all rel- evant information of its history. Nodes are labeled by instances of the productions that have lead to the completion of the case. Henthforth, they record the decisions (the choices among the allowed productions) together with information associated with these decisions. In the case of the editorial process, a terminated case contains the names of the referees, the evaluation reports, the editorial decision, etc. A terminated case is a tree whose branches reflect causal dependencies among subactivities used to solve a case, while abstracting from concurrent subactivities. The temporary information stored by the attributes attached to open nodes no longer exist when the case has reached completion. Closing nodes eliminates temporary information without resorting to any complex mechanism of distributed garbage collection. The artifacts can be collected in a log which may be used for the purpose of process mining [[27](#_bookmark23)] either for process discovery (by inferring a GAG from a set of artifacts using common patterns in their tree structure) or for conformance checking (by inspection of the logs produced during simulations of a model or executions of an actual implementation).

hal-00990007, version 1 - 13 May 2014

**Distribution.** Guarded attributed grammars can easily be implemented on a distributed ar- chitecture without complex communication mechanisms (shared memory, FIFO channels,...). Stakeholders in a business process own open nodes, and communicate asynchronously with other stakeholders via messages. Moreover there are no edition conflicts since each part of an artifact is edited by the unique owner of the corresponding node.

## Further works.

We plan to design prototypes to analyze and implement a GAG description together with the required support tools (editor, parser, checker, simulators ...) and to concentrate on the following research directions:

**Applicability.** We intend to develop some representative case studies to check applicability and limitations of the model. The first case study is a (simplified) distributed distance learning system, for which a GAGs implementation may have several advantages w.r.t. traditional solu- tions. First, since it does not rely on a client/server architecture, it should behave better in a degraded environment (when Internet connection is not always available). Most of the activity of the stakeholders are offline and the communication between them takes place upon avaibility of Internet connection. Second, the declarative decomposition of learning activities which does not impose a particular execution order together with the modularity of the model should provide more flexibility in the description of the learning processes. The second case study is a reporting system (e.g. semi-automatic synthesis of dashboards). The grammar can reflect the structure of the report, the identification of the stakeholders and their respective contributions lead to a distributed version of the grammar. Finally the semantic rules implement the automatic assem- bly of the reports. We obtain a “write things once” principle: once data are collected in some

RR n° 8528

30 *Badouel & al.*

synthesized position they can be used wherever they are needed. GAGs should hence reduce the workload of stakeholders: as the largest part of activity reports collect information that is already available somewhere, and can be extracted automatically by a GAG.

**Structuring the workspace of a stakeholder: Active Workspaces.** One can rely on the technique presented in [[26](#_bookmark23)] to extract a domain specific language from the subgrammar associated with a stakeholder according to the role(s) he plays in the system. The grammar may then contribute to structure the workspace of a stakeholder: Procedural parts of the artifacts given by the semantic rules encode and encapsulate technical know-hows that the end user may safely ignore. Each stakeholder manipulates documents in a familiar syntax using notations adapted to his domain of expertise (the DSL derived from the semantic rules). This can be important for enabling end-users with low level computer literacy to take part in the business process.

**Soundness.** Soundness is a crucial issue of case management systems: it guarantees that a case has a way to reach completion from any accessible configuration. Unsurprisingly soundness of GAGs in undecidable. We are looking for recursive subclasses of GAGs with decidable soundness.

hal-00990007, version 1 - 13 May 2014

# References

[1] Serge Abiteboul, Omar Benjelloun, Ioana Manolescu, Tova Milo, and Roger Weber. Active xml: A data-centric perspective on web services. In *BDA’02*, 2002.

[2] Kevin Backhouse. A functional semantics of attribute grammars. In *Tools and Algorithms for the Construction and Analysis of Systems, TACAS*, volume 2280 of *Lecture Notes in Computer Science*, pages 142–157. Springer, 2002.

[3] Laurian M. Chirica and David F. Martin. An order-algebraic definition of knuthian seman- tics. *Mathematical Systems Theory*, 13:1–27, 1979.

[4] Bruno Courcelle and Paul Franchi-Zannettacci. Attribute grammars and recursive program schemes i and ii. *Theor. Comput. Sci.*, 17:163–191 and 235–257, 1982.

[5] Elio Damaggio, Alin Deutsch, and Victor Vianu. Artifact systems with data dependencies and arithmetic. *ACM Trans. Database Syst.*, 37(3):22, 2012.

[6] Elio Damaggio, Richard Hull, and Roman Vaculín. On the equivalence of incremental and fixpoint semantics for business artifacts with guard-stage-milestone lifecycles. *Inf. Syst.*, 38(4):561–584, 2013.

[7] Pierre Deransart and Jan Maluszynski. Relating logic programs and attribute grammars.

*J. Log. Program.*, 2(2):119–155, 1985.

[8] Pierre Deransart and Jan Maluszynski. *A grammatical view of logic programming*. MIT Press, 1993.

[9] Clarence A. Ellis and Karim Keddara. Ml-dews: Modeling language to support dynamic evolution within workflow systems. *Computer Supported Cooperative Work*, 9(3/4):293–333, 2000.

[10] Harald Ganzinger. Increasing modularity and language-independency in automatically gen- erated compilers. *Sci. Comput. Program.*, 3(3):223–278, 1983.

Inria

*A Grammatical Approach to Data-centric Case Management* 31

[11] Harald Ganzinger and Robert Giegerich. Attribute coupled grammars. In *SIGPLAN Sym- posium on Compiler Construction*, pages 157–170. ACM, 1984.

[12] Richard Hull. Artifact-centric business process models: Brief survey of research results and challenges. In *OTM 2008*, volume 5332 of *Lecture Notes in Computer Science*, pages 1152–1163. Springer, 2008.

[13] Richard Hull, Elio Damaggio, Riccardo De Masellis, Fabiana Fournier, Manmohan Gupta, Fenno Terry Heath, Stacy Hobson, Mark H. Linehan, Sridhar Maradugu, Anil Nigam, Piyawadee Noi Sukaviriya, and Roman Vaculín. Business artifacts with guard-stage- milestone lifecycles: managing artifact interactions with conditions and events. In *Fifth ACM International Conference on Distributed Event-Based Systems, DEBS 2011*, pages 51–62. ACM, 2011.

[14] Thomas Johnsson. Attribute grammars as a functional programming paradigm. In *Func- tional Programming Languages and Computer Architecture, FPCA*, volume 274 of *Lecture Notes in Computer Science*, pages 154–173. Springer, 1987.

hal-00990007, version 1 - 13 May 2014

[15] Paul Klint, Ralf Lämmel, and Chris Verhoef. Toward an engineering discipline for gram- marware. *ACM Transaction on Software Engineering Methodologies*, 14(3):331–380, 2005.

[16] Donald E. Knuth. Semantics of context free languages. *Mathematical System Theory*, 2(2):127–145, 1968.

[17] Niels Lohmann and Karsten Wolf. Artifact-centric choreographies. In *Service-Oriented Com- puting - 8th International Conference, ICSOC 2010, San Francisco, CA, USA, December 7-10, 2010.*, pages 32–46, 2010.

[18] Alberto Martelli and Ugo Montanari. An efficient unification algorithm. *ACM Trans.*

*Program. Lang. Syst.*, 4(2):258–282, 1982.

[19] Brian H. Mayoh. Attribute grammars and mathematical semantics. *SIAM J. Comput.*, 10(3):503–518, 1981.

[20] Giorgio De Michelis and Clarence A. Ellis. Computer supported cooperative work and Petri nets. In *Advanced Course on Petri Nets, Dagstuhl 1996*, volume 1492 of *Lecture Notes in Computer Science*, pages 125–153. Springer, 1998.

[21] A. Nigam and N. S. Caswell. Business artifacts: An approach to operational specification.

*IBM Syst. J.*, 42:428–445, July 2003.

[22] OASIS. Web services business process execution language. Technical report, OASIS, 2007.

<http://docs.oasis-open.org/wsbpel/2.0/OS/wsbpel-v2.0-OS.pdf>.

[23] Jukka Paakki. Attribute grammar paradigms - a high-level methodology in language imple- mentation. *ACM Computing Surveys*, 27(2):196–255, 1995.

[24] João Saraiva and S. Doaitse Swierstra. Generating spreadsheet-like tools from strong at- tribute grammars. In *Generative Programming and Component Engineering, GPCE 2003*, volume 2830 of *Lecture Notes in Computer Science*, pages 307–323. Springer, 2003.

[25] João Saraiva, S. Doaitse Swierstra, and Matthijs F. Kuiper. Functional incremental attribute evaluation. In *Compiler Construction, CC 2000*, volume 1781 of *Lecture Notes in Computer Science*, pages 279–294. Springer, 2000.

RR n° 8528

32 *Badouel & al.*

[26] S. Doaitse Swierstra, Pablo R. Azero Alcocer, and João Saraiva. Designing and implementing combinator languages. In *Advanced Functional Programming*, pages 150–206, 1998.

[27] Wil M. P. van der Aalst. *Process Mining - Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

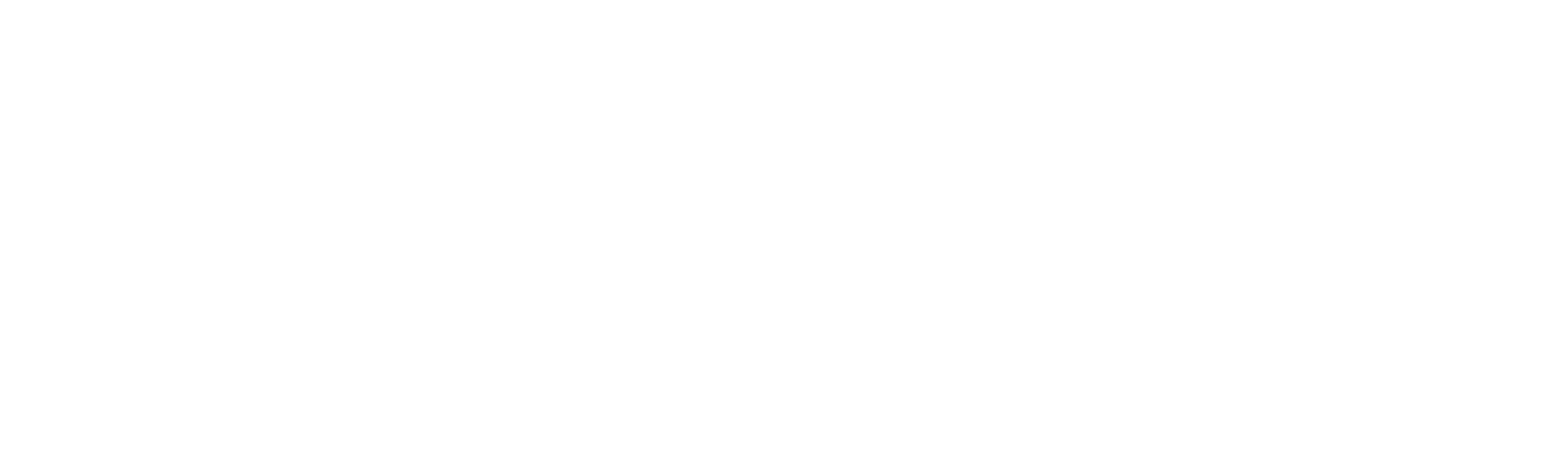
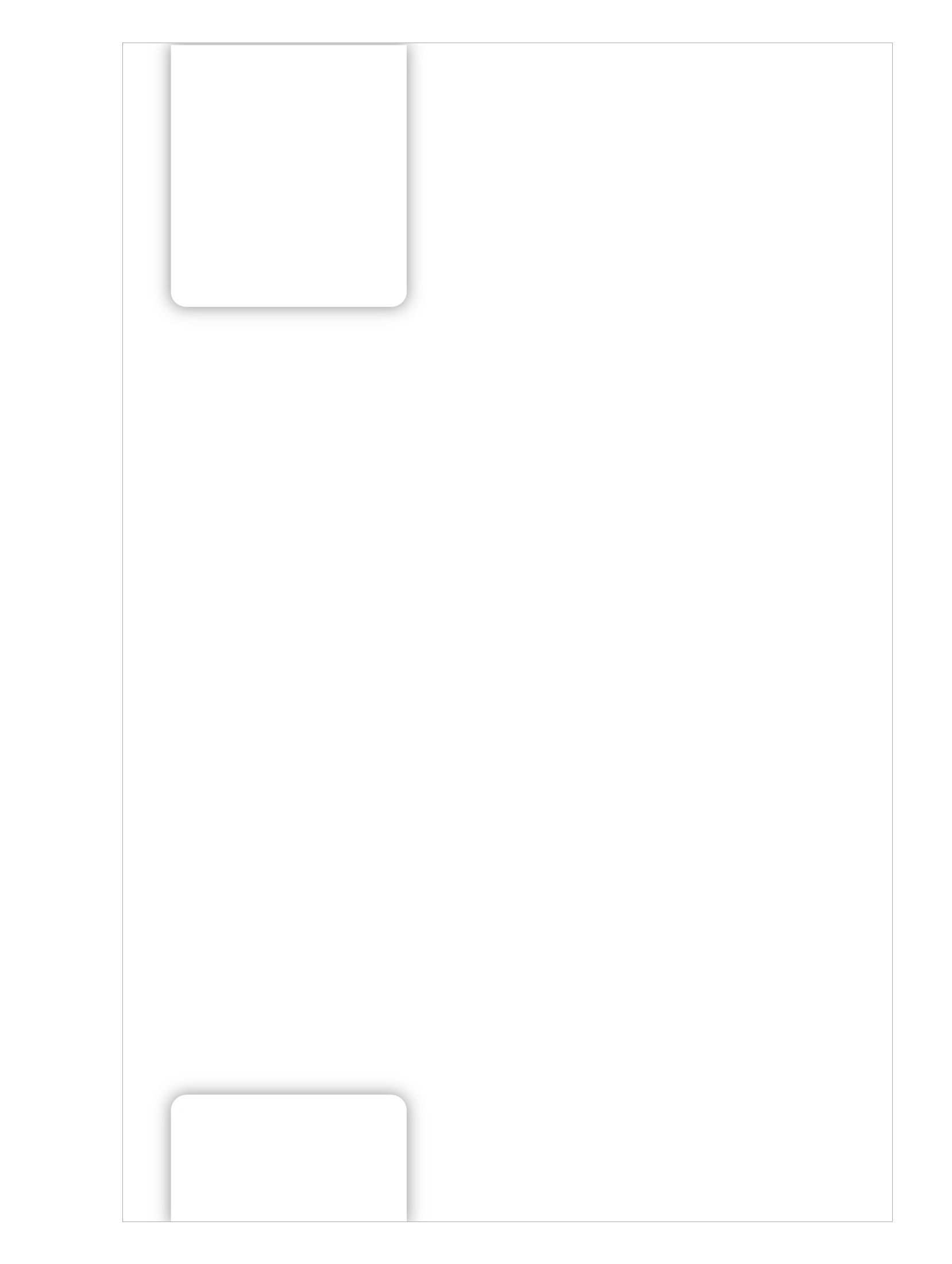
[28] Harald Vogt, S. Doaitse Swierstra, and Matthijs F. Kuiper. Higher-order attribute gram- mars. In *PLDI*, pages 131–145, 1989.

[29] Eric Van Wyk. Implementing aspect-oriented programming constructs as modular language extensions. *Sci. Comput. Program.*, 68(1):38–61, 2007.

[30] Eric Van Wyk, Oege de Moor, Kevin Backhouse, and Paul Kwiatkowski. Forwarding in attribute grammars for modular language design. In *Compiler Construction, ETAPS 2002, Grenoble, France*, pages 128–142, 2002.

hal-00990007, version 1 - 13 May 2014

Inria



hal-00990007, version 1 - 13 May 2014

**RESEARCH CENTRE**

**RENNES – BRETAGNE ATLANTIQUE**

Campus universitaire de Beaulieu 35042 Rennes Cedex

Publisher Inria

Domaine de Voluceau - Rocquencourt BP 105 - 78153 Le Chesnay Cedex inria.fr

ISSN 0249-6399