Application to NWO-EW Open Competition 2004, Second Round

1a Project Title: Graphs for Software Language Definitions
1b Project Acronym: GRASLAND
1c Principal Investigator: Dr. ir. A. Rensink, University of Twente

2a Summary. In the context of the MDA (Model Driven Architecture) methodology for designing maintainable software systems, model transformation is a central concept. Models are used to describe the system in all phases of development and on various levels of abstraction; they are specified in diverse (modeling and programming) software languages (SLs). Model transformations typically introduce concrete, implementation specific details.

Such transformations are intended to be correctness preserving: they should not introduce errors or essential changes. This, however, can be guaranteed only if the meaning of the SLs involved is defined with sufficient precision. Unfortunately, this is often lacking: many SLs have a well-defined syntax but only an informal semantics. A primary reason for this is that MDA does not include a general method for easily and consistently defining the subtler aspects of SLs, such as their semantics.

The purpose of this project is to define a meta-language in which all aspects of SLs, besides their concrete syntax, can be defined in a consistent manner. As a common formal foundation of this meta-language we propose graphs and graph transformations, which we believe to be powerful enough to capture all relevant SL aspects. This meta-language will enable us to provide semantic definitions of the source and target SLs involved in a given model transformation on a compatible basis; this in turn will enable us to precisely formulate and check the requirement of correctness preservation. We believe these abilities to be essential in realizing the full potential of MDA.

2b Lekensamenvatting. In de context van MDA (Model-Driven Architecture), een tamelijk recente methodologie voor het ontwerpen van goed onderhoudbare software-systemen waarvoor een groeiende belangstelling bestaat, neemt het transformeren van modellen een centrale plaats in. Het gaat hierbij om modellen die het te ontwikkelen software-systeem in een bepaalde ontwerfpase en op een bepaald abstractieniveau beschrijven; hun transformatie betreft bijvoorbeeld het toevoegen van concrete, implementatiespecifieke details.

De bedoeling is dat zulke modeltransformaties "correctheidbewarend" zijn, dat wil zeggen geen fouten of wezenlijke veranderingen in het gemodelleerde systeem introduceren. Zekerheid hieromtrent kan echter alleen bestaan als de betekenis van de gebruikte modeltalen voldoende duidelijk gedefinieerd is. Helaas ontbreekt het hier in de praktijk vaak aan: veel modeltalen hebben wel een goedgedefinieerde syntax maar slechts een informele semantiek. Als reden hiervoor is aan te voeren dat er in de context van MDA geen algemeen geaccepteerde methode bestaat om subtielere taalaspecten, zoals semantiek, op eenvoudige en consistente manier te definiëren.

Het doel van dit project is de definitie van een "meta-taal" voor het definiëren van softwarebeschrijvingstalen; een taal dus waarin (naast de concrete syntax) ook andere aspecten van model- en programmeertalen gedefinieerd kunnen worden. Aangezien de basis voor de semantiek van alle aldus gespecificeerde modeltalen hiermee gelijk komt te liggen, maakt dit het mogelijk om de eis dat modelttransformaties correctheidbewarend moeten zijn formeel vast te leggen, en deze eigenschap voor gegeven transformaties formeel te bewijzen. Dit is een noodzakelijke stap in het realiseren van de potentiële mogelijkheden van MDA.

3a) Informaticadisciplines:
3. Software Engineering: 3.2 (specificatiemethoden), 3.5 (interoperabiliteit), 3.6 (ontwikkeltools)
6. Fundamenten: 6.3 (semantiek), 6.5 (formele methoden)

3b) NOAG-i thema’s: Software Engineering (SE), Algorithms and Formal Methods (AFM)
4 Composition of the Research Team. The researchers that will be involved in this project are listed in the following table. Prof. dr. ir. M. Aksit will act as promotor of the OiO.

<table>
<thead>
<tr>
<th>Name (title)</th>
<th>Position</th>
<th>Expertise</th>
<th>Institute</th>
<th>fte</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. Aksit (prof.dr.ir.)</td>
<td>OiO (this project)</td>
<td>–</td>
<td>Univ. Twente</td>
<td>1.0</td>
</tr>
<tr>
<td>A. Rensink (dr.ir.)</td>
<td>HL Software Engineering</td>
<td>Software engineering</td>
<td>Univ. Twente</td>
<td>0.0</td>
</tr>
<tr>
<td>A. Kleppe (drs.)</td>
<td>Consultant</td>
<td>Graph transformation</td>
<td>Univ. Twente</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model-Driven Architecture</td>
<td>Klasse Objecten</td>
<td>0.1</td>
</tr>
</tbody>
</table>

5 Research School. The Software Engineering chair is a member of the research school: Instituut voor Programmatuurkunde en Algoritmiek (IPA).

6 Description of the Proposed Research

6.1 Context of the Proposed Research

A widely recognized proposal for combating the maintenance and evolution problems faced in software engineering is the **model driven approach**, brought to the world’s attention by the OMG’s Model Driven Architecture (MDA) framework [28]. MDA builds upon and greatly extends UML [46]; its cornerstones are meta-modeling and model transformation.

Key concepts in MDA are: **model**, **language**, **model transformation**, and **platform**. The relationships between these concepts are shown in Figure 1 (taken with permission from Fig. 8-8 of [28]).

Within MDA, a model is defined as a description of a software system written in a certain language, preferably a well-defined language (see Problem Description for our definition of a well-defined language). A model transformation converts one model into another model. Often the source

![Figure 1: Concepts and relationships in the MDA framework](image-url)
model is a so-called Platform Independent Model (PIM), whereas the target model is a Platform Specific Model (PSM). The concept of platform in this context can be described as the specific software and hardware that constitute the execution environment of the software system. Model transformations are automated and carried out by transformation tools. The input for the tools are transformation definitions.

Although MDA has received a warm welcome by the software developing community, currently the MDA vision cannot be implemented to its full extent, for the following reasons (among others):

1. UML, which is the language that is most often used for writing PIMs, lacks coherence and a well-defined semantics (see, e.g., [36, 22, 18, 12]).
2. The language in which model transformations should be written is not yet standardized. The OMG standardization process of what is now called the Query-View-Transformation (QVT, [34]) is still on its way, and the timing of its completion is uncertain.
3. The languages commonly used for writing PSMs, which are typically ordinary programming languages and extensions thereof, are usually not defined in a manner suited to build transformation definitions. Often a BNF grammar and a compiler are the only definitions provided.

This project addresses the third of these problems (and thereby implicitly part of the first, since UML is one of the SLs in question).

6.2 Problem Description

In the MDA Guide Version 1.0 ([28]) the concept of model transformation is defined as follows:

Model transformation is the process of converting one model to another model of the same system.

The fact that input and output models should describe the same system implies that model transformations are to preserve the meaning of the model(s) being transformed (see also [24]). For the meaning of the models we have to look at the semantics of the languages in which they are described. From now on we call these Software Languages (SLs), be it for implementation (i.e., programming languages) or specification.

The semantics of current SLs are often imprecise or even completely lacking (except through the “tool semantics” provided by the compiler). Therefore it is difficult to determine with any degree of rigor whether a given model transformation really meets the above criterion. We hold that for building useful model transformations, it is imperative that the SLs in which the source and target model are written are well-defined. We call a language well-defined when the following aspects are clear and unambiguous:

- The concrete syntax: i.e., the notation used for writing models;
- The abstract syntax: i.e., the concepts available for constructing models;
- The static semantics: i.e., the elements and conceptual relationships of the systems described by the models;
- The dynamic semantics: i.e., the behavior of the systems described by the models, as can be examined by the changes between snapshots taken at subsequent points in time;
- The relationships between these four aspects.
As remarked above, many SLs used in the context of MDA lack one of more of these aspects. Especially the requirement that static and dynamic semantics are closely related and should not be defined separately is important for SLs. Software is by definition dynamic. Even simple data storage systems have an important dynamic side (the so-called CRUD functions). A definition of an SL therefore cannot be complete without a definition of the dynamic semantics.

The established manner for defining concrete syntax is (E)BNF. The OMG standard called Meta-Object Facility [33] has been proposed as a means for defining abstract syntax. What is lacking, however, is a general way to define the relationship between concrete and abstract syntax, for instance along the lines of attribute grammars. Even less clear in the context of MDA is a convenient means for consistently defining static and (operational) dynamic semantics, in a way that is compatible with the abstract syntax — even though the essential theory has been known for some time (e.g., [48, 31]).

Note that it is not even necessarily the case that there is only one possible semantics to a given (concrete and abstract) syntax. A case in point are UML statecharts, for which a large number of semantics definitions have been published (e.g., [27]). Due to this plethora, if the intended semantics is not explicitly given it can still be unclear whether a given system built from a statechart indeed conforms to its intended specification.

What MDA needs in order to fulfill its promise is a special-purpose language definition language (LDL) for defining all four SL aspects in a compatible framework; in other words, a meta-language in which multiple SLs can be defined so as to uphold the well-definedness requirements stated above. The goal of this project is the definition of such a meta-language.

6.3 Approach

In this section we propose a solution strategy to the problem described above, based on two key principles. (We use \( L \) to denote an arbitrary given SL and \( p \) for a “program” [model, instance] of \( L \).)

1. We explicitly recognize, for each of the SL aspects \( \text{Asp} \) identified above, on the one hand the meta-model \( \text{Asp}_L \) defining that aspect for \( L \), and on the other hand the process of constructing model instances \( \text{Asp}(p) \) of \( \text{Asp}_L \) for concrete programs \( p \) of \( L \).

2. We propose to use graphs and graph grammars to define the models, meta-models (interpreted as type graphs) and transformation processes for all SL aspects.

**Example.** We demonstrate these principles using a running example. We define a toy Software Language XL, consisting just of label declarations and goto statements, with the following grammar:

\[
\text{Prog} ::= \text{Stat} ; ' \text{Prog}' \mid \text{Stat} \ documentation \ .
\]

\[
\text{Stat} ::= \text{Label} \mid \text{GoTo} .
\]

\[
\text{Label} ::= \text{Id} .
\]

\[
\text{GoTo} ::= \text{go} \text{Id} .
\]

At a statement of the form \( \text{go} \ x \), the system jumps non-deterministically to any of the places where the label \( x \) is declared. The following is our running example program \( xp \) in this language:

\[
a ; \ \text{go} \ a ; \ a .
\]

**Concrete Syntax.** The concrete syntax of \( L \) is typically given in EBNF, which is a string grammar. The grammar gives rise to a parser for \( L \) which can construct a syntax tree \( \text{CoSy}(p) \) from syntactically...
correct instances $p$ of $L$. The structure of the syntax trees is implicit in the rules of the grammar: the nodes are terminals and non-terminals, and the branching structure is derived from the placement of the corresponding textual fragments within the original $p$.

For the purpose of this project we require an explicit specification of the meta-model $\text{CoSy}_L$ of the concrete syntax trees of $L$, in addition to the rules of the grammar; moreover, the grammar is taken as a special case of a graph grammar.

**Example.** The first column of Figure 2 shows the concrete syntax meta-model $\text{CoSy}_{\text{XL}}$ and the concrete syntax tree $\text{CoSy}(\text{xp})$. Note that we have made the ordering of the children in the syntax tree explicit.

**Abstract Syntax.** As stated in the previous section, in the context of MDA the abstract syntax for a given SL $L$ is typically defined through a (possibly annotated) diagrammatic meta-model; in the setting of this project, this corresponds to a type graph $\text{AbSy}_L$. Instances $\text{AbSy}(p)$ (for $p$ a program of $L$) are to be derived from the corresponding concrete syntax tree $\text{CoSy}(p)$; the derivation can be defined again through a graph transformation from $\text{CoSy}_L$-typed graphs to $\text{AbSy}_L$-typed graphs. What happens during this transformation is obviously specific to $L$, but typically it will involve the merging and removal of redundant syntactic elements, as well as name resolution.

**Example.** The abstract syntax of $\text{XL}$ rearranges the statement types in an abstraction hierarchy and unifies all label identifiers. The resulting type graph $\text{AbSy}_{\text{XL}}$ as well as the example instance $\text{AbSy}(\text{xp})$

---

3What we describe may be called a concrete syntax tree; it is actually more common to work with so-called abstract syntax trees, in which those syntactic elements that were merely there to ease the job of the parser have been removed. This is in fact a step towards our abstract syntax graphs described below. To avoid confusion we concentrate on concrete syntax trees in this proposal.

---

**Figure 2:** Example meta-models and models for the program $\text{xp}$: “$a$; $\text{go } a$; $a$.”
are shown in the second column of Figure 2. The transformation from the concrete to the abstract syntax involves removing the intermediate Stat-nodes (while maintaining the nxt-ordering) and merging Id-nodes with identical names.

**Static Semantics.** The static semantics of a program is typically expressed using a collection of models, including, for instance, flow, data, call and dependency graphs. In this project we take the position that these can fruitfully be combined into a single model, possibly at the loss of ready visualization but certainly at the gain of consistency. The separate models can be recaptured as views upon this combined model. Again, the static semantics of a language \( L \) can be specified through a meta-model, \( \text{StSe}_L \); instances can be derived from the abstract syntax graphs through graph transformation.

**Example.** The static semantics of \( XL \) adds Begin- and End-nodes and control flow to the abstract syntax graph, as mentioned above. The meta-model \( \text{StSe}_{XL} \) and example model \( \text{StSe}(xp) \) are given in the third column of Figure 2. The transformation rule for a Label-statement just converts the outgoing nxt-edge into a flow-edge; the rule for a GoTo-statement creates a flow-edge to all Label-statements defining the label identifier used by the GoTo.

**Dynamic Semantics.** In contrast to the aspects discussed so far, the dynamics of a software system will be described not by one single model, but by a collection of snapshots with a transition relation between them. Each of these snapshots can be modeled as an extension of the static semantics model with run-time information; in language implementation terms, the run-time information consists of (abstractions of) the stack and heap of a program. Together this gives rise to a state-transition system, which can be seen as the behavioral model of the system.

Once more, we can capture the structure of the snapshots in a meta-model, \( \text{DySe}_L \). It will generally extend the static semantics (\( \text{StSe}_L \)) with the types of run-time information mentioned above. Transformation rules in this context play a different role than before: rather than collectively describing the construction of one model for a given program \( p \) from another (concrete to abstract syntax to static semantics), now each rule will describe a (small) step in the execution of \( p \); in other words, each rule acts as a graph-based virtual machine instruction.

**Example.** Figure 3 shows the meta-model \( \text{DySe}_{XL} \) (which adds a node type VM for the “virtual machine” executing a given program and an edge type pc for the “program counter” pointing to the statement under execution) as well as the transition system modeling the behavior of \( xp \).

---

**Figure 3: Dynamic semantics of the example program**
Meta-language As discussed and illustrated above, the key idea is to describe all essential elements of SL definitions using (type) graphs, and the relationship between these elements using graph transformations. Thus, graphs and graph transformations are the core of the meta-language (the LDL) which this project is out to define. We rely on the extensive research in graph transformations to provide us with results on simulation and verification and algorithms for automatic parsing, synthesis and transformation of models. This establishes a common basis for a more precise definition of language aspects, and thereby a foundation on which a theory of correctness-preserving, intra- or inter-language model transformations can be built — thus addressing, and hopefully closing, one of the main gaps in the road towards realizing the benefits of the MDA approach.

6.4 Project Objectives

We define the following steps/milestones in the project:

1. A case study consisting of a programming language definition, addressing all aspects mentioned above, for a realistic fragment of an existing programming language, such as Java. (Preliminary work in this direction has been done in two MSc theses: see [8, 23].)
2. A case study consisting of a specification language definition, addressing all aspects mentioned above; for instance, UML state diagrams. The intention here is not to define a new semantics but to formulate one of the existing semantic definitions in terms of graphs and transformations, along the lines described above; see, e.g., [47] for the feasibility of such a semantics.
3. An initial definition of an LDL in which at least the language concepts identified in the case studies can be specified. This requires an understanding of these concepts on an abstract enough level. At this stage, especially the abstract syntax and semantics of the LDL are of interest.
4. A concrete syntax for the LDL, once the initial definition (previous step) is stable.
5. In parallel with the steps above and partly driving them, the (re)formulation of the case study language definitions (steps 1 and 2) in terms of the LDL being developed.
6. The development of tools for compiling LDL “programs” (i.e., SL specifications), which means the creation of the meta-models and graph transformation systems.
7. The complete definition of the case study languages (steps 1 and 2) in the LDL.

All in all, the outcome of the project is a language definition language with formal semantics and (light-weight) tool support, and the formal definition of a realistic programming and a specification language in the LDL.

6.5 Other Related Work

This project has clear connections with work in graph rewriting and other approaches to the specification and verification of behavioral semantics, as well as the development of formal foundations for UML and MDA.

Graph rewriting. An authoritative and comprehensive overview of the theory and practice of graph rewriting can be found in [45, 15]. Graph rewriting has been applied before in models of computation; see, e.g., [7, 16, 14, 11]. Our own project GROOVE [20, 4] investigates the combination of graph rewriting and verification techniques (in particular, model checking).

Within software engineering there are several currently very active areas of application of graph rewriting, such as software evolution (e.g., [30]) and UML model and meta-model semantics (e.g., [19, 21, 26, 47]).
Behavioral semantics Both process algebra ([10]) and programming language theory (e.g., [48]) have developed a rich theory of behavioral semantics, from which we intend to benefit in this project. In fact the graph transformation semantics we intend to pursue is obviously a form of operational semantics, and it seems natural to incorporate a form of SOS (Structural Operational Semantics, see [6]) into our LDL. Another connection exists to term graph rewriting [9], which is indeed a technique for graph transformation-based operational semantics of functional languages.

Formal foundations for UML. There are many research groups investigating semantics for UML; e.g., the aforementioned statechart semantics [27] and graph rewriting approaches [26], but also the work carried out in the pUML group [17, 25, 37]. The difference with this project is that we are not interested in any particular semantics, but with the means to define them systematically for a given SL.

6.6 Embedding within own existing activities

Software Engineering chair. Within the software engineering group, a number of topics are being studied, some pertaining to various stages of the software development process, others to the process itself. The principal themes are (i) Aspect-Oriented Software Composition, (ii) Synthesis Based Software Architecture Design, (iii) Soft Computing applied to Software Development, and (iv) Design for Correctness. The current proposal falls into (iv) and (to a lesser degree) in (ii).

Klasse Objecten. The core activities of Klasse Objecten are consultancy and training in software design through modeling, using UML, OCL and MDA. From 1997 Klasse Objecten has been actively involved in creating the UML and OCL standards. The knowledge gained by this participation is crucial for the company. In the coming years the subject of MDA will become more and more important to (clients of) Klasse Objecten. Therefore it is necessary for Klasse Objecten to be involved in the development of stable foundations for the application of MDA.

Principal investigator. Arend Rensink moved to the area of graph transformation in 2001, after having done research in process algebra-related fields for close to a decade. His particular interest is the application of graphs as software models and graph transformations to describe the dynamics of those models; this raises opportunities for software modeling and verification that are new in the context of the (otherwise quite well-established) area of graph transformations. Apart from a number of conference publications (see reference section), the research has resulted in the development of a tool environment called GROOVE [38], to be developed further in an NWO project of the same name [20] started recently. The current project dovetails onto GROOVE very nicely, since the LDL to be developed here will allow to provide graph transformation semantics to SLs that can then be fed into the GROOVE tool.

7 Work Programme

Planning. The tasks described below refer to the enumeration in the project objectives (Section 6.4).

<table>
<thead>
<tr>
<th>Tasks</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 2</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 3</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>— Consolidation and PhD thesis —</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Education. It is anticipated that the OiO takes part in the regular IPA courses. Depending on the background of the OiO, he will also participate in a selection of (international) Master courses.
9 Literature

Key publications of the research team


Other references


[37] The precise uml group. See http://www.cs.york.ac.uk/puml/.


