Little Languages: Little Maintenance?

ARIE VAN DEURSEN
CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands; arie@cwi.nl

PAUL KLINT
CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands; paulk@cwi.nl and
University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

ABSTRACT

So-called *little*, or *domain-specific* languages (DSLs), have the potential to make software maintenance simpler: domain-experts can directly use the DSL to make required routine modifications. On the negative side, however, more substantial changes may become more difficult: such changes may involve altering the domain-specific language. This will require compiler technology knowledge, which not every commercial enterprise has easily available. Based on experience taken from industrial practice, we discuss the role of DSLs in software maintenance, the dangers introduced by using them, and techniques for controlling the risks involved.

1 Introduction

Little languages, tailored towards the specific needs of a particular domain, can significantly ease building software systems for that domain (Bentley, 1986). To cite Herndon and Berzins (1988),

> If a conceptual framework is rich enough and program tasks within the framework
> are common enough, a language supporting the primitive concepts of the framework
> is called for. (...) Many tasks can be easily described by agreeing upon an appropriate
> vocabulary and conceptual framework. These frameworks may allow a description
> of a few lines long to replace many thousand lines of code in other languages.

We will use the following terminology (see also Figure 1):

Domain-Specific Language (DSL) A small, usually declarative, language expressive over the distinguishing characteristics of a set of programs in a particular problem domain (Walton, 1996).
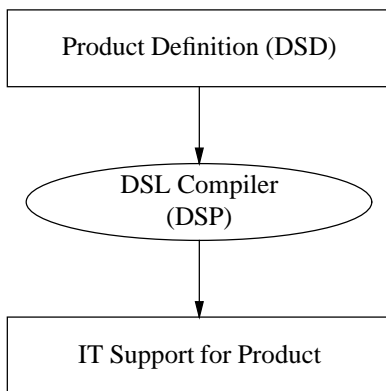
```
┌─────────────────────────────────┐
│     Product Definition (DSD)     │
└─────────────────────────────────┘
                  │
                  ▼
         ╭─────────────────╮
         │   DSL Compiler   │
         │     (DSP)        │
         ╰─────────────────╯
                  │
                  ▼
┌─────────────────────────────────┐
│     IT Support for Product       │
└─────────────────────────────────┘
```

Figure 1: A DSL compiler.

Domain-Specific Description (DSD) A "program" (specification, description, query, process, task, ...) written in a DSL.

Domain-Specific Processor (DSP) A software tool for compiling, interpreting, or analyzing domain-specific descriptions.

A well-designed DSL will help the application builder to write short, descriptive, and platform-independent DSDs. Moreover, the good DSL will be effectively implementable, where the DSPs capture the stable concepts and algorithmic ingredients of the particular domain. Using such a DSL for constructing domain-specific applications, increases reliability and repairability, provides self-documenting and portable descriptions, and reduces forward (and backward) engineering costs (Herndon and Berzins, 1988). Over the past few years many different languages have been developed, such as Yacc for describing grammars and parsers (Johnson, 1975), PIC for drawing pictures (Kernighan, 1982), PRL5 for maintaining database integrity within digital switching systems (Ladd and Ramming, 1994), Morpheus for specifying network protocols (Abbot and Peterson, 1993), and GAL for specifying video device drivers (Thibault *et al.*, 1997).

In this paper, we elaborate on the advantages and problems of the use of domain-specific languages, emphasizing their role in software maintenance. Evidently, the attributes listed above will help reduce maintenance costs, and for that reason domain-specific approaches are investigated in order to arrive at "inherently maintainable software" (Glover and Bennet, 1996). However, using a domain-specific language can also make a system more difficult to maintain, for example if changes to the underlying domain model become necessary.

To discuss these issues, we first give an example of the commercial use of a DSL taken from the area of financial engineering (Section 2). We then cover the implications for software maintenance, and identify the risks and opportunities involved in the use of a DSL (Section 3). We conclude by describing two techniques (Sections 4 and 5) that will help to address two of the potential problems in the use of DSLs.

2 The Financial Engineering Domain

2.1 Interest Rate Products

Financial engineering deals, amongst others, with *interest rate products*. Such products are typically used for inter-bank trade, or to finance company take-overs involving triple comma figures in multiple currencies. Crucial for such transactions are the protection against and the well-timed exploitation of risks coming with interest rate or currency exchange rate fluctuations.

The simplest interest rate product is the loan: a fixed amount in a certain currency is borrowed for a fixed period at a given interest rate. More complicated products, such as the *financial future*, the *forward rate agreement*, or the *capped floater* (Coggan, 1995, Chapter 12), all aim at *risk reallocation*. Banks can invent new ways to do this, giving rise to more and more interest rate products. Not surprisingly, different interest rate products have much in common, making financial engineering an area suitable for incorporating domain-specific knowledge in tools, languages, or libraries. As an example, Eggenschwiler and Gamma (1992) describe the ET++ SwapsManager, an object-oriented framework for manipulating interest rate products. We refer to van Deursen (1997) for a comparison of domain-specific languages and object-oriented frameworks in the financial engineering domain.

2.2 Challenges

A software system supporting the use of interest rate products typically deals with the bank's financial administration (who is buying what), and — more importantly — provides management information allowing decision makers to assess risks involved in the products currently processed. Typical problems found in such systems are that it is:

- too difficult to introduce a new type of product, even if it is very similar to existing ones;

- impossible to ensure that the instructions given by the financial engineer are correctly implemented by the software engineer.

The first problem leads to a long time-to-market for new products; the second leads to potentially incorrect behavior.

2.3 The Risla Language

Dutch bank MeesPierson, together with software house CAP Volmac saw the use of a specific language for describing interest rate products as the solution to the problems of long time-to-market and potentially inaccurate implementations. The language was to be readable for financial engineers, and descriptions in this language were to be compiled into COBOL. In this section we summarize earlier (and more detailed) accounts given by van Deursen (1994); Arnold *et al.* (1995); van den Brand *et al.* (1996b) of the development and use of this language.

The development of this language, called RISLA (for Rente Informatie Systeem Language — Interest rate information system language), started in 1992, and can be summarized as follows:

- MeesPierson had a very good library of COBOL routines for operating on cash flows, intervals, interest payment schemes, date manipulations, etc.
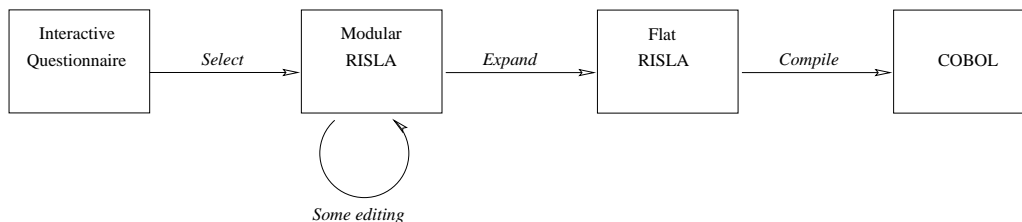
Figure 2: From questionnaire, via Modular and flat Risla, to COBOL

- Using this library directly in COBOL did not provide the right level of abstraction, and cumbersome encoding tricks were needed to use, e.g., lists without a fixed length.

- An interest rate product can be considered as a "class": it contains instance variables to be assigned at creation time (the principal amount, the interest rate, the currency, etc.), information methods for inspecting actual products (when is interest to be paid), and registration methods for recording state changes (pay one redemption).

The language RISLA was designed to describe interest rate products along these lines. An instantiated product is called a *contract*, fixing the actual amount, rate, etc. of a particular product sold. The language is based on a number of built-in data types for representing cash flows, intervals, etc., and has a large number of built-in operations manipulating these data types (the operations correspond to the subroutines in the COBOL library). A product definition specifies the contract parameters, information methods, and registration methods.

RISLA is translated into COBOL. Other systems in the bank can invoke the generated COBOL code to create new contracts, to ask information about existing contracts, or to update contract information. The initial version of RISLA was used to define about 30 interest rate products.

After a few years of working with RISLA, the users found out that the modularization features of RISLA were inadequate. A RISLA description defines a complete product; but different products are constructed from similar *components*. To remedy this situation, a project *Modular* RISLA was started. RISLA was extended with a small modular layer, featuring parameterization and renaming. Moreover, a *component library* was developed, and the most important products were described using this library.

In addition to that, the RISLA development team made an effort to make the language more accessible to the financial experts. To that end, an inter-active *questionnaire* interface to the component library was developed. End-users can combine existing components into a new product by filling in the answers of a questionnaire.

This use of questionnaires and modular RISLA gives rise to the financial product life cycle as shown in Figure 2. An interactive questionnaire is filled in, and the answers are used to select the relevant RISLA components. This definition may contain some holes that are specific to this product, which can be filled by writing the appropriate RISLA code. The modular definition is then expanded to a flat (non-modular) definition, which in turn is compiled into COBOL.

As a last point of interest, the actual questionnaire used is defined using a second domain-specific language: RISQUEST. This is a language for defining questions together with permitted answers (choice from a fixed set, free text). Moreover, RISQUEST has constructs for indicating in which order questions are to be asked, and how this sequencing may depend on the actual

answers given. Last but not least, RISQUEST can be used to associate library components with the possible answers. A RISQUEST definition is entered in textual form, and it is generated into a Tcl/Tk program. This program can be invoked by a financial engineer to fill in the questionnaire and to generate the corresponding modular RISLA.

2.4 Evaluation

On the positive side, the RISLA project has met its targets: the time it costs to introduce a new product is down from an estimated three months to two or three weeks. Moreover, financial engineers themselves can use the questionnaire to compose new products. Last but not least, it has become much easier to validate the correctness of the software realization of the interest rate products.

On the negative side, it is not so easy to extend the language. When a new data type or a new built-in function is required, the compiler as well as the COBOL library, need to be adapted. This requires skills in compiler construction technology, which is not the typical background of people working mainly in a COBOL environment. Finally, the RISLA product definitions have become longer and longer. Whenever there was a new software system requiring information about products that was not provided in the existing methods, new methods had to be provided, sometimes requiring new data types or extensions to the RISLA language.

3 The Maintenance Perspective

3.1 Benefits of DSLs

The single most important benefit of using domain-specific languages is that the domain-specific knowledge is formalized at the right level of abstraction. This, in turn, has the advantages that:

- Domain experts themselves can understand, validate, and modify the software by adapting the domain-specific descriptions (DSDs).

- Modifications are easier to make and their impact is easier to understand.

- Domain-specific knowledge is explicitly available, and not hidden into, e.g., COBOL code. As a result the use of a DSL avoids the need for *business rule extraction*. In fact, a DSL can be the output of such a reverse engineering activity.

- The explicitly available knowledge can be re-used across different applications.

- The way the knowledge is represented is independent of the implementation platform; the DSPs hide whether the DSDs are translated into C, Fortran, COBOL, ...

Concerning the costs of using DSLs, there is empirical evidence suggesting that the use of DSLs increases flexibility, productivity, reliability, and usability (Kieburtz *et al.*, 1996). In the context of the FAST approach developed at Bell Labs, a productivity increase with a factor of four to five has been reported (Weiss, 1997). Within this approach, small languages (called *jargons*) can be defined for representing information dealing with, e.g., employees, recipes, email-messages, etc. Files containing descriptions written in jargons can then be processed by a series

of language-specific tools called *wizers*. A wizer is written in a syntax-directed manner using pattern matching and can take advantage of a well-developed library of functions typically required when writing wizers (Nakatani and Jones, 1997).

As a way of reducing the costs of the initial development of the DSL and DSPs, the language and its tools can be sold as a product to competitors in the same field. In this way, it is possible to earn back initial development costs but at the same time keeping secret the suite of DSDs (DSL programs) that describe the company's proprietary products.

3.2   DSL Development

The development of a DSL requires a thorough understanding of the underlying domain. The steps to be taken include (see also (Cleaveland, 1988)):

- Identify problem domain of interest.

- Gather all relevant knowledge in this domain.

- Cluster this knowledge in a handful of semantic notions and operations on them.

- Construct a library that implements the semantic notions.

- Design a DSL that concisely describes applications in the domain.

- Design and implement a compiler that translates DSL programs to a sequence of library calls.

- Write DSL programs for all desired applications and compile them.

The first two steps raise the important question how to recognize a domain, and how to determine the scope of a domain. Following Simos (1995, 1996), two definitions of *domain* could be used. The first is generally used in the artificial intelligence and object-oriented communities. It lets a domain correspond to the "real world", without direct relation to software systems it might be encoded in. The second definition comes from the systematic software reuse research community. It defines a domain as "a set of systems including common functionality in a specified area".

In principle, a DSL could be designed following either definition. Most benefits in terms of reducing maintenance costs, however, are to be expected from the "domain as a set of systems" approach. Following Simos (1996), candidate domains should be (1) mature, i.e., a set of legacy systems exists; (2) reasonably stable, i.e., certain aspects of these systems are satisfactory and worth studying; and (3) economically viable, i.e., new systems are anticipated in the domain. The pieces of functionality in the legacy systems will help to identify the domain. Furthermore, modification requests from the past, or differences between the various systems, will help to identify the *variability* in the domain. The DSL should then be designed such that it is expressive over this variability.

When collecting information about a domain, techniques from the *domain analysis* (Arango, 1989; Lam and Whittle, 1996) or *domain engineering* (Simos, 1995, 1996) methodologies will be of help. These will also aid in determining the scope of the domain at hand. The Organization Domain Modelling (ODM) approach to domain engineering (Simos, 1996), for example, views

domain demarcation as determining a *shared set of boundary agreements*. ODM proposes a number of steps to arrive at domain boundaries that meet certain criteria, such as being explainable, shareable, and evolvable.

3.3   DSL Design Questions

With respect to software maintenance, there are a number of considerations to be taken into account during the design of a DSL:

- Who will be writing the DSDs? What is the expected domain-specific background, and how much programming knowledge is required?

- How many DSDs will there be needed, and how long will they be? It may be possible to validate the correctness of three pages of DSL code, but who is going to predict the impact of a change in one out of 100 DSDs, each 25 pages long?

- Which (decidable) forms of static analysis and which integrity checks on DSDs are anticipated?

- What should happen if it turns out that the language requires new data types or new functionality?

  One approach could be to give the DSL sufficient expressive power to define new data types or data operations, but this complicates the construction of the DSPs. For example, some form of iteration or recursion increases expressive power, but making the language Turing complete will make the verification of important properties (e.g., termination) undecidable.

  The users of RISLA decided to keep the language as simple as possible. In order to reduce compiler maintenance when new functions in the language are needed, they decided to parameterize the compiler with an easily extensible table mapping RISLA functions to COBOL procedures.

- Does the DSL support user-definable syntax for, e.g., naming procedures? This may increase the readability, an important issue in DSLs, but it seriously complicates the construction of DSPs, including analysis tools that are needed during later maintenance phases.

- Is the main library written in the DSL or written in the target language? Who will be responsible for maintaining the library?

- Is the interface (data representation) to other systems easily adaptable or is it hidden inside the implementation of the DSL compiler?

- Who will be responsible for maintaining the DSPs? Is the knowledge about the domain sufficiently stable such that changes in the design of the DSL or the DSP are not to be expected?

The actual trade-off to be made for each of these issues clearly depends on the domain and the application at hand, and on the prominence maintenance considerations take during the DSL design.

#	Maintainability Factor	DSL Influence
1.	Ease of expressing anticipated modifications	++
2.	Small development costs per application	++
3.	Small code size (low LOC)	++
4.	Low annual change traffic (ACT)	0
5.	Code readability	++
6.	System modularity	++
7.	Locality of changes	++
8.	Testability	+
9.	Code portability	+
10.	Maintenance process followed	+
11.	Maintainability as an objective	+
12.	Quality of configuration management	0
13.	Repository for modification requests	0
14.	Small number of languages used	−−

Figure 3: Maintainability factors, and the best possible effect (ranging from negative $--$ via neutral 0 to positive $++$) the use of a DSL has on each of these factors.

3.4   Maintainability Factors

Define high maintainability as low expected costs per modification request. Figure 3 lists a series of "maintainability factors" affecting these costs, taken mostly from (Pigoski, 1997; Peercy, 1996; Oman and Hagemeister, 1994). For each of these factors, we have indicated what the best possible effect on it will be from using a DSL. Observe that:

- The maintenance costs M for a given application are often expressed as (Boehm, 1981):

$$M = F * D * ACT$$

  where D is the initial development costs of that application, F is some weight factor depending on, e.g., the type of system, the language used, experience of the maintainers team, and *ACT* is the *Annual Change Traffic*, the fraction of code changed due to maintenance.

  When using a DSL, *D* will decrease significantly (factor # 1. in Figure 3 has label "++"), *F* may decrease, and *ACT* will remain the same (factor # 4. has label "0"), when compared to a system developed in a general purpose language. Here we assume a linear relationship between the number of lines needed in the general purpose and in the domain-specific language to implement a typical modification request.

- Using a DSL, the end-user (the domain expert) can be involved in the maintenance process: he or she may be able to express the modification request in terms of the DSL, or to validate the correctness of the modifications made. For this reason, factors # 10 and # 11 have been given a "+" label.

3.5 Risks

The maintenance risks involved in the use of DSLs can partly be related to making the wrong trade-offs in the design questions listed before. Other issues include:

- The use of a DSL involves a shift from maintaining hand-built applications towards maintaining (1) DSDs (DSL programs defining each application); (2) DSPs (the DSL compiler); (3) a DSL library of predefined objects. Especially maintaining the DSL compiler requires skills not available in every organization.

- For existing, widely used, languages one can profit from readily available manuals, tutorials, courses, and experienced people. For a new DSL one has to develop this all from scratch.

- Any language needs a minimal number of users in order to survive: It may be the case that there will be too few users of the particular DSL.

- An organization that is active in various domains, may need a large number of DSLs. This can be advantageous, in that it will help the organization to build up routine in language development. Care has to be taken, however, to minimize the differences between the various languages.

- For related, but different, application areas different DSLs are needed. How can applications based on them cooperate?

In the remaining sections, we will discuss two techniques to alleviate two (the first and the last) of these risks. In this paper, we stress the use of these techniques for designing and coordinating DSLs. We refer to van den Brand *et al.* (1996a) for a discussion of the same techniques emphasizing their application to reengineering and system renovation problems.

4 Designing and Implementing DSLs

As mentioned above, the use of a DSL has important benefits, but moves part of the maintenance problems to the DSP level. In this section, we discuss the ASF+SDF Meta-Environment, and how it supports the development and maintenance of application languages. It was in fact used during the design of RISLA and RISQUEST, the languages described in Section 2.

It is the aim of ASF+SDF to assist during the design and further development of (domain-specific) languages (Bergstra *et al.*, 1989; Klint, 1993; van Deursen *et al.*, 1996). It consists of a formalism to describe languages and of a Meta-Environment to derive tools from such language descriptions. Ingredients often found in an ASF+SDF language definition include the description of the (1) context-free grammar, (2) context-sensitive requirements, (3) transformations or optimizations that are possible, (4) operational semantics expressing how to execute a program, and (5) translation to the desired target language. The Meta-Environment turns these into a parser, type checker, optimizer, interpreter, and compiler, respectively.
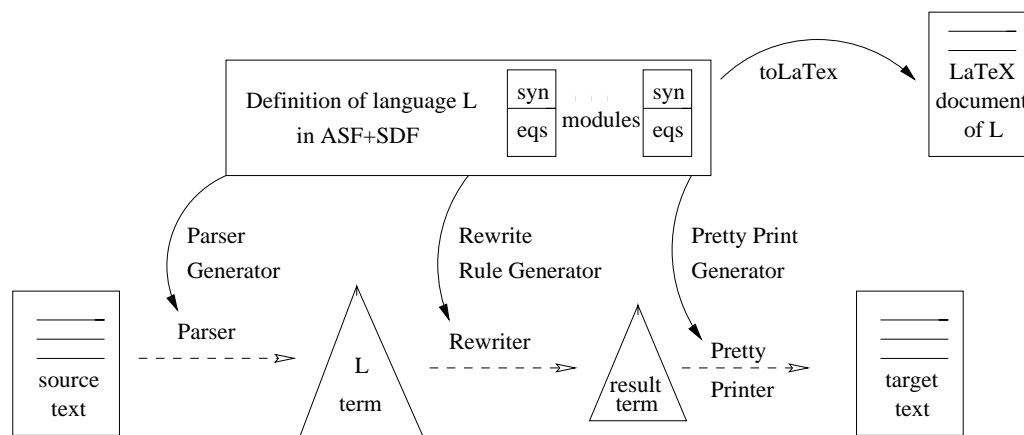
Figure 4: A language definition for L in the ASF+SDF Meta-Environment.

4.1 The ASF+SDF Formalism

The language ASF+SDF grew out of the integration of the Algebraic Specification Formalism ASF and the Syntax Definition Formalism SDF (Bergstra *et al.*, 1989). An ASF+SDF specification consists of a declaration of the functions that can be used to build *terms*, and of a set of equations expressing *equalities* between terms.

If we use ASF+SDF to define a language L, the grammar is described by a series of functions for constructing abstract syntax trees. Transformations, type checking, translations to a target language L', etc., are all described as functions mapping L to, respectively, L, Boolean values, and L'. These functions are specified using conditional equations, which may have negative premises. In addition to that, ASF+SDF supports so-called *default*-equations, which can be used to "cover all remaining cases", a feature which can result in significantly shorter specifications for real-life situations (van Deursen *et al.*, 1996). Specification in the large is supported by some basic modularization constructs.

Terms can be written in arbitrary user-defined syntax. In fact, an ASF+SDF signature is at the same time a context-free grammar, and defines a fixed mapping between sentences over the grammar and terms over the signature. Thus, an ASF+SDF definition of a set of language constructors specifies the concrete as well as the abstract syntax at the same time. Moreover, concrete syntax can be used in equations when specifying language properties. This smooth integration of concrete syntax with equations is one of the factors that makes ASF+SDF attractive for language definition.

4.2 The ASF+SDF Meta-Environment

The role of the ASF+SDF Meta-Environment (Klint, 1993) is to support the development of language definitions, and to produce prototype tools from these. It is best explained using Figure 4. A modular definition of language L, generates *parsers*, which can map L-programs to L-terms, *rewriters*, which compute functions over L-programs by reducing terms to their normal form, and *pretty printers*, which map the result to a textual representation. In the Meta-Environment,

the generators are invisible, and run automatically when needed. The derived pretty printer can be fine-tuned, allowing one to specify compilers to languages in which layout is semantically relevant, such as COBOL (van den Brand and Visser, 1996).

This pattern gives rise to a series of language processors, with a functionality as specified in the language definition. Basic user-interface primitives can be used to connect the processors to an integrated L-specific environment.

The ToLaTeX facility of the ASF+SDF Meta-Environment encourages the language designer to write his or her definition as a *literate* specification.

4.3   Industrial Applications

The typical industrial usage of ASF+SDF is to build tools for the analysis and transformation of programs in existing languages as well as for the design and prototyping of domain-specific languages. In this paper we will concentrate on the latter. The ASF+SDF formalism is used to write a formal language definition, and the Meta-Environment is used to obtain prototype tools. Once the language design is stable and completed successfully, the prototype tools can — depending on the needs of the application — be re-implemented in an efficient language like C, although there are also examples in which the generated prototype is satisfactory, and re-implementation is not even considered.

The underlying observation is that language design is both critical and difficult, and that it should not be disturbed by implementation efforts in a language like C. At the same time, prototype tools are required during the design phase to get feedback from language users. ASF+SDF helps to obtain these tools with minimal effort, by executing the language definitions, and by offering a number of generation facilities.

This requires an extra investment during the design phase, since ASF+SDF enforces users to write a thorough language definition. The assumption is that this investment will pay for itself during the implementation phase, an assumption confirmed by the various projects carried out so far, such as the ones discussed by van den Brand *et al.* (1996b).

5   Coordinating different DSLs

So far we have seen how language technology can be applied to design and prototype a specific DSL and how to build supporting tools for DSL programs. In general, however, one will need a whole range of DSLs to cover the application areas that occur in a large organization. How can applications that have been built by means of different DSLs be coordinated? We answer this question in two steps: first we introduce the TOOLBUS coordination architecture and then we show how it solves the coordination issue just raised.

5.1   The TOOLBUS coordination architecture

Bergstra and Klint (1996b,a) have proposed the TOOLBUS coordination architecture facilitating the interoperability of heterogeneous, distributed, software components. To get control over the possible interactions between components ("tools") direct inter-tool communication is forbidden. Instead, all interactions are controlled by a "**T** script" that formalizes all the desired interactions among tools. This leads to a communication architecture resembling a hardware communication bus.
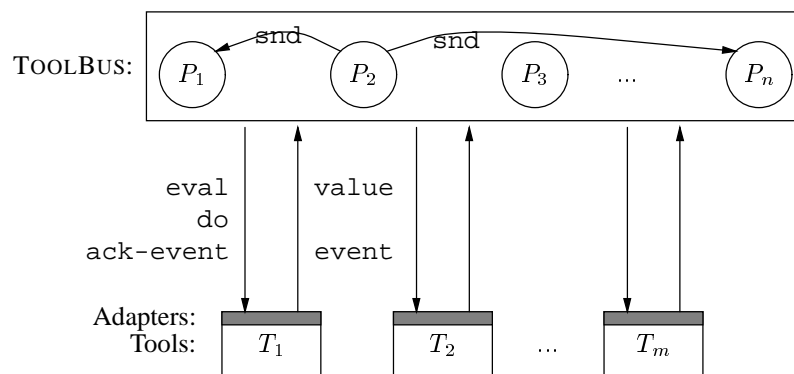
Figure 5: Global organization of the TOOLBUS

The global architecture of the TOOLBUS is shown in Figure 5. The TOOLBUS serves the purpose of defining the cooperation of a variable number of *tools* T_i ($i = 1, ..., m$) that are to be combined into a complete system. The internal behavior or implementation of each tool is irrelevant: they may be implemented in different programming languages, be generated from specifications, etc. Tools may, or may not, maintain their own internal state. Here we concentrate on the external behavior of each tool. In general an *adapter* will be needed for each tool to adapt it to the common data representation and message protocols imposed by the TOOLBUS.

The TOOLBUS itself consists of a variable number of processes P_i ($i = 1, ..., n$). The parallel composition of the processes P_i represents the intended behavior of the whole system. Tools are external, computational activities, most likely corresponding with operating system level processes. They come into existence either by an execution command issued by the TOOLBUS or their execution is initiated externally, in which case an explicit connect command has to be performed by the TOOLBUS. Although a one-to-one correspondence between tools and processes seems simple and desirable, this is not enforced and tools are permitted that are being controlled by more than one process as well as clusters of tools being controlled by a single process.

At the implementation level, the **T** script is executed by an interpreter that makes connections with tools via TCP/IP. In various case studies tools for user-interfacing, data storage and retrieval, parsing, compiling, constraint solving, scheduling, simulation and game-playing have been successfully integrated in various combinations yielding seamlessly integrated applications although the building blocks used are heterogeneous and may even execute in a distributed fashion.

5.2   Exchanging data

When coordinating distributed, heterogeneous, components, two key questions should be answered:

- How do components exchange data?

- How is the flow of control between components organized?

The former is discussed here, the latter is postponed to Section 5.3. There are two alternatives for exchanging data between components. One can either provide a direct mapping between the

machine/language-specific representations of data in the various components or one can provide a common representation to which all machine/language-specific representations are converted.

In the case of the TOOLBUS the latter approach has been chosen and simple prefix terms are used as common data representation. Terms may consist of integers, strings, reals, function applications (e.g., `f(1,2)`) and lists (e.g., `[1, "abc", 3]`). For most applications this suffices, but as a general escape mechanism, terms may contain so-called binary strings that can represent arbitrary binary data such as, for instance, object files and bitmaps.

At the implementation level, terms are compressed before they are shipped between components, thus enabling fast exchange of large amounts of data.

5.3 T scripts

A **T** script describes the overall behavior of a system and consists of a number of definitions for processes and tools and one TOOLBUS configuration describing the initial configuration of the system. A process is defined by a process expression, and a tool by the name of its executable. Process behavior is based on a variant of Discrete Time Process Algebra and provides primitives for

- synchronous, binary, communication ("messages");

- asynchronous, broadcasting communication ("notes");

- tool-related actions such as creation/connection, communication, and termination/disconnection;

- process composition operators such as sequential composition, choice, iteration, parallel composition, and conditional;

- remote monitoring of processes and tools;

- delay and timeout.

5.4 Examples

A typical application of the TOOLBUS approach is shown in Figure 6. From the user's perspective, a database management system (DBMS) can be queried through a graphical user-interface (GUI). From an architectural perspective, the GUI and the DBMS are completely decoupled and they are even running on different machines. The key issue here is that there is no fixed connection between the components; both only communicate with the TOOLBUS and the processes running there (e.g., P_1 and P_2) determine the routing of GUI requests to the DBMS. This is achieved using the various communication primitives available in **T** scripts. The routing may even be changed dynamically, without disturbing the overall operation of the application.

Other examples are a distributed auction (where one auction master and a variable number of bidders cooperate in an auction, each working via his/her own workstation), distributed multi-user games, multi-user distributed programming environments and the like.

In all these examples, the **T** script defines the global architecture of each application and a wide variety of components based on a range of implementation technologies can be fitted into this architecture provided that they obey the protocol imposed by the **T** script.
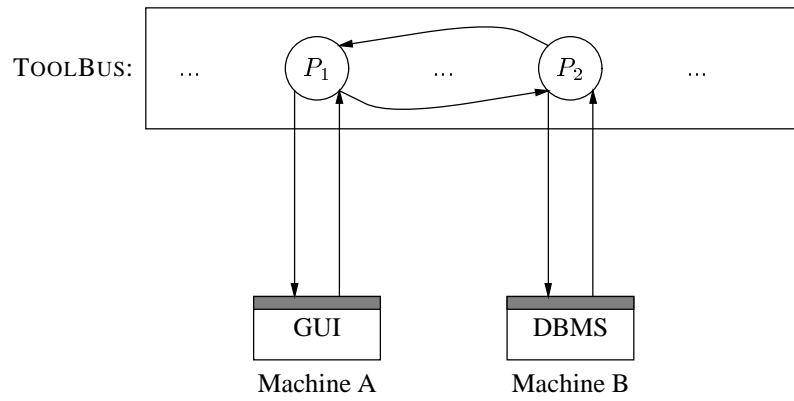
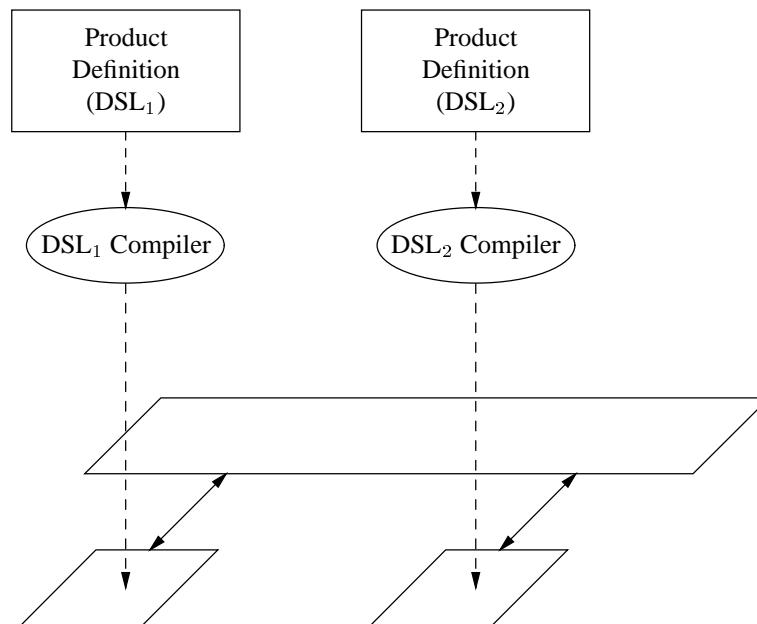Figure 6: A Typical distributed application.

Figure 7: Coordination of DSLs using the TOOLBUS.

Further examples and a discussion of related work are out of the scope of this paper, but provided by Bergstra and Klint (1996a,b). The TOOLBUS is currently in use in various projects inside and outside our own research group. We are using it ourselves as the basic communication infrastructure for a reimplementation of the ASF+SDF Meta-Environment (van den Brand *et al.*, 1997). As a result, tools like parsers, editors, compilers for *different* languages can be integrated in a unified framework. This is, in particular, due to the use of a common data format to exchange information between tools (e.g. parse trees).

5.5 Coordinating DSLs with the TOOLBUS

Applications that have been constructed by means of different DSLs can be coordinated using the TOOLBUS technology as well. Recall from Figure 1 the case of a product definition in some DSL and its compilation to the desired IT support for that product. In Figure 7 we sketch the case where two different products are being defined using two different DSLs and how they can be coordinated. Typically, all DSL compilers will generate TOOLBUS compatible components and an overall script will describe the cooperation of all (generated) components.

There are several issues involved here related to maintenance, renovation, and gradual evolution:

- The TOOLBUS acts as a form of "middleware" that can connect new and old software components. It enables the gradual transition from a system based on traditional, hand-crafted, components to a system based on generated components using DSLs.

- Maintenance of a specific DSL or its compiler does not affect the whole system.

- Different DSLs can use different technology (when relevant). This enables transitions to new technology during the evolution of a system.

- For flexibility and ease of maintenance, each DSL compiler can also be based on a private TOOLBUS (not shown in Figure 7).

6   Concluding remarks

6.1   Assessment

DSLs are no panacea for solving all software engineering problems, but a DSL designed for a well-chosen domain and implemented with adequate tools may drastically reduce the costs for building new applications as well as for maintaining existing ones.

On the positive side, in a DSL-based approach one concentrates all knowledge about an application in the DSL and its supporting component libraries, while all implementation knowledge is concentrated in the DSP (DSL compiler). From the perspectives of flexibility, quality assurance, maintenance, and knowledge management this is a highly desirable situation.

On the negative side, an application domain may not yet be sufficiently understood to warrant the design of a DSL for it or adequate technology may not be available to support the design and implementation of the DSL. Under such circumstances a more traditional approach to system design and maintenance should be preferred.

An alternative to a DSL is using a general purpose (object-oriented) language together with a library of data types and functions for the domain in question. Software development based on DSLs (following the steps from Section 3.2) is a way to encourage software engineers to indeed build such a library, to provide the most natural notation (the DSL) for accessing the library, and to think in advance of the typical changes to be expected in that particular domain.

6.2   Future directions

We have already mentioned that the usability of DSLs by application domain experts (as opposed to programmers) is a decisive factor for their acceptance and success. There are several directions for increasing the ease of use of DSLs:

- Visual DSLs in which visual/iconic user-interfaces are used to compose library components.

- Natural language DSLs in which stylized natural language sentences are used to compose applications.

- Interactive DSLs in which domain experts are guided through a list of queries in order to select and assemble an application from library components.

- Prototyping environments for DSLs that support the realistic simulation of applications.

Regarding the design and implementation of DSLs we see the following needs:

- Further development of tools for designing and implementing DSLs. Typical issues: (1) modular structure of the DSL; (2) static checking of DSDs (DSL programs); (3) correctness of the translation rules used by the DSP.

- Tools for designing and implementing supporting component libraries. Typical issues: (1) modular structure and design of the component library; (2) implementation of the modular structure in given implementation languages, e.g., how to implement parameterized modules in COBOL? There is a relation here with current work on designing so-called *business objects*.

- Tools for connecting different DSLs. Typical issue: while coordination architectures as described in Section 5 provide basic connectivity and interoperability, a more abstract, application level, model of coordination is needed.

- Collection of empirical data concerning maintenance costs in systems built using domain-specific languages, following, e.g., the maintenance metrics as proposed by Grady (1987).

Domain specific languages ("little languages") introduce an appropriate abstraction level for packaging domain-specific knowledge and technology. "Little maintenance" is becoming feasible for applications using them, provided that state-of-the-art techniques are used like the ones discussed in this paper.

References

Abbot, M. B. and Peterson, L. L. (1993). A language-based approach to protocol implementation. *IEEE/ACM Transactions on Networking*, **1**(1).

Arango, G. (1989). Domain analysis: From art form to engineering discipline. In *Fifth International Workshop on Software Specification and Design*, pages 152–159. Appeared as ACM SIGSOFT Engineneering Notes 14(3).

Arnold, B. R. T., van Deursen, A., and Res, M. (1995). An algebraic specification of a language for describing financial products. In M. Wirsing, editor, *ICSE-17 Workshop on Formal Methods Application in Software Engineering*, pages 6–13. IEEE.

Bentley, J. L. (1986). Programming pearls: Little languages. *Communications of the ACM*, **29**(8), 711–721.

Bergstra, J. A. and Klint, P. (1996a). The Discrete Time ToolBus. In M. Wirsing and M. Nivat, editors, *Algebraic Methodology and Software Technology (AMAST '96)*, volume 1101 of *Lecture Notes in Computer Science*, pages 288–305. Springer-Verlag.

Bergstra, J. A. and Klint, P. (1996b). The ToolBus coordination architecture. In P. Ciancarini and C. Hankin, editors, *Coordination Languages and Models (COORDINATION '96)*, volume 1061 of *Lecture Notes in Computer Science*, pages 75–88. Springer-Verlag.

Bergstra, J. A., Heering, J., and Klint, P., editors (1989). *Algebraic Specification*. ACM Press/Addison-Wesley.

Boehm, B. W. (1981). *Software Engineering Economics*. Prentice-Hall.

van den Brand, M. G. J. and Visser, E. (1996). Generation of formatters for context-free languages. *ACM Transactions on Software Engineering and Methodology*, **5**, 1–41.

van den Brand, M. G. J., Klint, P., and Verhoef, C. (1996a). Core technologies for system renovation. In K.G.Jeffrey, J. Král, and M. Bartošek, editors, *SOFSEM '96: Theory and Practice of Informatics*, LNCS, pages 235–254. Springer-Verlag.

van den Brand, M. G. J., van Deursen, A., Klint, P., Klusener, S., and van der Meulen, E. A. (1996b). Industrial applications of ASF+SDF. In M. Wirsing and M. Nivat, editors, *Algebraic Methodology and Software Technology (AMAST '96)*, volume 1101 of *Lecture Notes in Computer Science*, pages 9–18. Springer-Verlag.

van den Brand, M. G. J., Olivier, P., Moonen, L., and Kuipers, T. (1997). Implementation of a prototype of the new ASF+SDF Meta-Environment. In A. Sellink, editor, *Theory and Practice of Algebraic Specifications, ASF+SDF'97*, Electronic Workshops in Computing. Springer-Verlag. To appear.

Cleaveland, J. C. (1988). Building application generators. *IEEE Software*, pages 25–33.

Coggan, P. (1995). *The Money Machine: How the City Works*. Pinguin. Third edition.

van Deursen, A. (1994). *Executable Language Definitions: Case Studies and Origin Tracking Techniques*. Ph.D. thesis, University of Amsterdam.

van Deursen, A. (1997). Domain-specific languages versus object-oriented frameworks: A financial engineering case study. In *Proceedings Smalltalk and Java in Industry and Academia, STJA'97*, pages 35–39, Erfurt. Ilmenau Technical University.

van Deursen, A., Heering, J., and Klint, P., editors (1996). *Language Prototyping: An Algebraic Specification Approach*, volume 5 of *AMAST Series in Computing*. World Scientific Publishing Co.

Eggenschwiler, T. and Gamma, E. (1992). ET++ SwapsManager: Using object technology in the financial engineering domain. In *OOPSLA'92*, pages 166–177. ACM. SIGPLAN Notices 27(10).

Glover, S. J. and Bennet, K. H. (1996). An agent-based approach to rapid software evolution based on a domain model. In *Proceedings International Conference on Software Maintenance ICSM'96*, pages 228–237. IEEE Computer Society Press. Monterey, CA.

Grady, R. B. (1987). Measuring and managing software maintenance. *IEEE Software*, pages 35–45.

Herndon, R. M. and Berzins, V. A. (1988). The realizable benefits of a language prototyping language. *IEEE Transactions on Software Engineering*, **SE-14**, 803–809.

Johnson, S. C. (1975). Yacc – yet another compiler compiler. Technical Report 32, AT&T Bell Laboratories.

Kernighan, B. W. (1982). PIC — a language for typesetting graphics. *Software — Practice and Experience*, **12**(1), 1–21.

Kieburtz, R. B., McKinney, L., Bell, J. M., Hook, J., Kotov, A., Lewis, J., Oliva, D. P., Sheard, T., Smith, I., and Walton, L. (1996). A software engineering experiment in software component generation. In *Proceedings of the 18th International Conference on Software Engineering ICSE-18*, pages 542–553. IEEE.

Klint, P. (1993). A meta-environment for generating programming environments. *ACM Transactions on Software Engineering and Methodology*, **2**, 176–201.

Ladd, D. A. and Ramming, J. C. (1994). Two application languages in software production. In *USENIX Very High Level Languages Symposium Proceedings*, pages 169–178.

Lam, W. and Whittle, B. (1996). A taxonomy of domain-specific reuse problems and their solutions. *ACM SIGSOFT Software Engineering Notes*, **21**(5), 72–77.

Nakatani, L. and Jones, M. (1997). Jargons and infocentrism. In *Proceedings of the first ACM SIGPLAN Workshop on Domain-Specific Languages*, pages 59–74.

Oman, P. and Hagemeister, J. (1994). Constructing and testing of polynomials predicting software maintainability. *Journal of Systems and Software*, **24**(3), 251–266.

Peercy, D. (1996). Improving the maintainability of software – book review. *Journal of Software Maintenance*, **8**, 345–356.

Pigoski, T. M. (1997). *Practical Software Maintenance – Best Practices for Managing Your Software Investment*. John Wiley and Sons.

Simos, M. (1995). Organization domain modeling (ODM): Formalizing the core domain modeling life cycle. In M. Samadzeh and M. Zand, editors, *Proceedings of the Symposium on Software Reusability SSR'95*, pages 196–205. ACM Software Engineering Notes.

Simos, M. (1996). Organization domain modelling (ODM) guidebook version 2.0. Technical Report STARS-VC-A025/001/00, Synquiry Technologies, Inc. URL: `http://www.organon.com/`. 450 pp.

Thibault, S., Marlet, R., and Consel, C. (1997). A domain-specific language for video device drivers: From design to implementation. In *Proceedings of the USENIX Conference on Domain-Specific Languages*, pages 11–26, Berkeley, CA.

Walton, L. (1996). Domain-specific design languages. URL http://www.cse.ogi.edu/-∼walton/dsdls.html.

Weiss, D. (1997). Creating domain-specific languages: The FAST process. In S. Kamin, editor, *First ACM-SIGPLAN Workshop on Domain-Specific Languages; DSL'97*. Technical report, University of Illinois, Department of Computer Science. See URL at http://www-sal.cs.uiuc.edu/∼kamin/dsl.