Declarative Specification of Bidirectional Transformations Using Design Patterns

KEVIN LANO1, SHEKOUFEH KOLAHDOUZ-RAHIMI2, AND SOBHAN YASSIPOUR-TEHRANI1

1Department of Informatics, King’s College London, London WC2B 4BG, U.K.
2Department of Software Engineering, University of Isfahan, Isfahan 81746-73441, Iran
Corresponding author: Kevin Lano (kevin.lano@kcl.ac.uk)

ABSTRACT Bidirectional transformations (bx) are a specific form of model transformation (MT) used in model-driven engineering to maintain consistency between two models, which may change independently. Currently bx are defined using a number of specialized transformation languages, which have had limited uptake due to complex semantics and poor efficiency. In contrast, unidirectional transformation languages such as ATL have been widely adopted, but require separate forward and reverse transformations to be written to address model synchronization requirements. In this paper, we provide declarative specification techniques for bx, systematically constructed using MT design patterns. We define two approaches to declarative bx definition: 1) by automatically bidirectionalizing unidirectional transformation specifications and 2) by developing specification guidelines for the QVT-R standard language to make it more effective for bx in practice. The approaches are evaluated using a large-scale code-generator bx from UML to ANSI C and other examples. Their semantic validity is demonstrated by rigorous arguments.

INDEX TERMS Bidirectional transformations, design patterns, model transformations, QVT-R, UML-RSDS.

I. INTRODUCTION
Model-driven engineering (MDE) and Model-based development (MBD) have reached a point of significant application in industry, particularly in the automotive industry and more recently in finance. These development approaches use models (such as UML class diagrams) as key artifacts within their processes. In some cases the models are used as the complete definition of an application, with executable code being generated automatically from the models.

Model transformations (MT) are a key element of MDE and MBD approaches for software, supporting code generation, migration, refinement, reverse-engineering and many other capabilities. Bidirectional model transformations (bx) are a specific type of MT which provide a means to maintain a consistency relation R between two models which may both change. This is in contrast to unidirectional transformations, which establish R by constructing a target model from a source model, and do not subsequently maintain the consistency of these models.

Bx are significant for a wide range of model based engineering scenarios involving model synchronization:
- Incremental execution of transformations, for example, it is desirable for a code generator to only make the minimal necessary changes to generated code if a source model is changed, rather than regenerating the entire code set again.
- Round-trip engineering [28], where either an application specification or design may change and the consistency of the two models at different levels of abstraction must be automatically maintained.
- Automatic consistency maintenance of multiple inter-related models at the same level of abstraction [18], such as UML class diagrams and sequence diagrams, to support agility in model-based development.
- Application and database co-evolution [5].

The motivating case for this paper was a code generator from UML to ANSI C [44]. For realistic applications involving large-scale models with thousands of elements, incremental code generation of C is a practical necessity. Additionally, the use of a bx in this case is motivated by considerations of semantic correctness of the generated code, and by the desirability of round-trip engineering capabilities.

A bx can be used to enforce a relation R between two different representations of similar semantic information. For example, a UML class and a corresponding ANSI C struct. Such a bx is responsible for ensuring that changes to one of these representations is propagated to the other to maintain their mutual consistency. There may however be additional
information in one model which has no counterpart in the other. For example, in the well-known Families to Persons bx benchmark [1], the persons model includes birthday information, whilst the families model differentiates parents and children. These features are specific to just one of the linked models.

In the theoretical framework of [59], bx are characterized by a binary relation \( R : SL \leftrightarrow TL \) between a source language (metamodel) \( SL \) and a target language TL. \( R(m, n) \) holds for a pair of models \( m \) of \( SL \) and \( n \) of TL when the models consist of data which correspond under \( R \).

The key idea of a bx is that instead of defining two separate transformations from \( SL \) to \( TL \) and from \( TL \) to \( SL \), to establish \( R \), it should be possible to automatically derive from the definition of \( R \) both forward and reverse transformations

\[
R^\rightarrow : SL \times TL \rightarrow TL
\]

\[
R^\leftarrow : SL \times TL \rightarrow SL
\]

which establish/maintain \( R \) between the two models, given both existing source and target models. This eliminates the risk that forward and reverse transformations could become inconsistent with each other, and reduces the workload involved in maintaining two separate transformation specifications instead of one.

The \( R^\rightarrow, R^\leftarrow \) transformations should satisfy two key conditions [46], [59]:

1) Correctness: the forward and reverse transformations derived from a relation \( R \) do establish \( R \):

\[
R(m, R^\rightarrow(m, n)) \quad R(R^\leftarrow(m, n), n)
\]

for each \( m : SL, n : TL \).

2) Hippocraticness: if source and target models already satisfy \( R \) then the forward and reverse transformations do not modify the models [48]:

\[
R(m, n) \implies R^\rightarrow(m, n) = n
\]

\[
R(m, n) \implies R^\leftarrow(m, n) = m
\]

for each \( m : SL, n : TL \).

Hippocraticness is a global property, in practice a stronger local property is desirable: if any part of a target (source) model is already consistent with the corresponding part of the source (target) model, then neither part should be modified by forward or reverse transformations. We refer to this as local Hippocraticness. Further, it is preferable that updates to a model by \( R^\rightarrow \) or \( R^\leftarrow \) should be minimal necessary to restore \( R \): this is termed the principle of least change [48].

In the following, we will consider only separate-models transformations, where source and target models are distinct, and not update-in-place transformations, which operate on a single model which is both the source and target.

We describe two alternative approaches to bx definition:

1) In the first approach, unidirectional separate-models transformations \( \tau \) are defined declaratively using the Object Constraint Language (OCL) [52], and a bx relation \( R \) and forward and reverse transformations \( R^\rightarrow, R^\leftarrow \) are derived syntactically from the specification of \( \tau \).

We consider specifically the unidirectional MT language UML-RSDS [42].

2) In the second approach, a bx \( \tau \) is defined using a restricted subset of QVT-R [51], with the specification of \( \tau \) organized so that both \( R^\rightarrow \) and \( R^\leftarrow \) exist and satisfy the bx properties wrt \( R \) derived from the specification.

An original contribution of our work is an emphasis on the role of design patterns in the definition of \( \tau \) in order to achieve bx properties.

Section II describes related work, Section III describes UML-RSDS, Section IV describes model transformation and bx design patterns, and Section V defines our approach to bx specification in UML-RSDS. Section VI considers inverse transformations and the bx properties of \( R^\rightarrow \). Section VII describes a large case study of a bx transformation between UML and C. Section VIII describes our QVT-R based approach to bx definition. Section IX gives an evaluation of the approaches, and Section X gives conclusions. The appendix provides an argument that the bx properties for \( R^\rightarrow \) are satisfied by transformations defined using our first approach.

II. RELATED WORK

Bx are the subject of extensive theoretical research [10], [12], [46], [58]–[60] and an annual conference, BX. The most well-known transformation language which supports definition of bx is QVT-R [51], which uses an OCL-based expression language. However, QVT-R has had limited uptake because of its complex semantics and specification style, compared to simpler unidirectional MT languages such as ATL [31] and ETL [33]. The semantic basis of the language remains unclear [8], [46], [60]. The combination of bx and update-in-place processing involves aspects such as implicit deletion (if an element is not required to exist by any rule, then it is deleted), which are difficult to use, and are implemented differently by different QVT-R tools [8], [40]. Although the language is intended to be declarative, the meaning of a specification can depend on procedural aspects (rule invocation relationships) in addition to logical predicates. As the examples of [46] show, this leads to specifications which are difficult to understand and liable to subtle errors.

Triple Graph Grammars (TGG) are a graph transformation-based formalism which supports bx definition [14], although with a restricted expression language compared to OCL. Other approaches, such as JTL [11], and extensions of TGG [2], use constraint-based programming techniques to interpret relations \( P(s, t) \) between source and target elements as specifications in both forward and reverse directions. Model finding using Alloy is used to bidirectionalize ATL in [46], and to backward propagate view changes in [58]. These approaches have problems of efficiency for large-scale models. An alternative constraint-based approach
is to generate constraints restricting target models, instead of the models themselves [13]. This allows greater freedom of choice in the construction of the target model, but requires an additional step to actually derive specific models.

Another research direction has focused on the special case of asymmetric bx known as lenses, in which the source to target mapping is independent of the existing target model. The target model \( f : TL \) is computed as a view or abstraction \( \text{get}(s) \) containing partial information from the source model \( s : SL \) [6], [15]. A dual map \( \text{put} : SL \times TL \rightarrow SL \) propagates view changes to the source, such that \( \text{put}(s, \text{get}(s)) = s \) and \( \text{get}(\text{put}(s, t)) = t \). Various semantic and syntactic techniques have been developed for deriving the put function from get in different cases [34], [61].

In previous papers we have briefly introduced some bx design patterns [39] and techniques for bx verification [41]. In this paper we extend these previous works by comprehensively defining an approach to bx specification using UML-RSDS, and we identify additional bx patterns. We also introduce an approach using QVT-R with the bx patterns. We provide rigorous arguments to show the correctness of the UML-RSDS approach, provide an evaluation, and give new examples and extensive details of a large case study.

III. UML-RSDS

In our first approach to bx definition, we look at a unidirectional MT language, UML-RSDS. In this approach, unidirectional transformations \( \tau \) are specified using UML use cases with OCL postconditions \( Post_\tau \) and invariants \( Inv_\tau \), then a statically-computed bx relation \( R_\tau \) for \( \tau \) is defined as \( Post_\tau \land Inv_\tau \) for \( \tau \), together with a forward transformation \( \tau^{-1} \) extending \( \tau \), and an inverse \( \tau^{-1} \) of \( \tau^{-1} \). These play the role of \( R^{-1} \) and \( R^{-1} \) in bx theory. The bx relation and transformations are derived from \( \tau \) using a higher-order transformation. This is generally more efficient than the use of constraint programming and model finding.

UML-RSDS is a precise subset of UML 2 and OCL 2.4, with a formal semantics [35] and an established toolset [36], [42]. We use a variant notation for OCL, with \( E.ALLInstances() \) abbreviated to \( E \) for entity types on the lhs of a \( \rightarrow \) operator. \( x : s \) is used for \( s \rightarrow \text{includes}(x) \), \( s1 <: s2 \) for \( s2 \rightarrow \text{includesAll}(s1) \), & is used for conjunction, \( \implies \) for implication. \(^1\) Logical expressions \( P \) may be used operationally, to stand for behaviors \( \text{stat}_L(C)(P) \) which establish \( P \) (Section III-C). In particular, the predicate \( x \rightarrow \text{isDeleted}() \) is used to express that the element \( x \) is to be removed from all entity types and other collections in which it resides. \( s \rightarrow \text{isDeleted}() \) for collection \( s \) means \( s \rightarrow \text{forall}(x \rightarrow \text{isDeleted}()) \). The general iterate operator is excluded from our subset, as are tuples and invalid. null cannot be explicitly referred to, but only tested using \( x \rightarrow \text{oclIsUndefined}() \).

\(^1\) Using symbols for logical connectives helps to visually distinguish these from the parts of the formulae that they connect.

A. TRANSFORMATION SPECIFICATION

Model transformations \( \tau \) are specified in UML-RSDS as UML use cases, defined declaratively by two main predicates, expressed in our OCL variant:

1) Postconditions \( Post_\tau \) which define the intended effect of the transformation at its termination. These are an ordered conjunction \( C_1 \land \ldots \land C_n \) of OCL constraints (also termed rules in the following) and also serve to define a procedural implementation of the transformation.

2) Invariants \( Inv_\tau \) which define expected invariant properties which should hold during the transformation execution (their truth is preserved by individual transformation rule applications). The invariants can be derived from \( Post_\tau \), or can be specified explicitly by the developer.

From a declarative viewpoint, \( Post_\tau \) defines the conditions which should be established by the transformation. From an implementation perspective, the constraints of \( Post_\tau \) also define the intended computation steps of the transformation: each computation step is an application of a postcondition constraint (transformation rule) to a specific source model element or to a tuple of source model elements. The computation steps should preserve \( Inv_\tau \).

A typical postcondition constraint/rule has the form:

\[
S_i ::
Cond \implies Pred
\]

where \( S_i \) is the context entity type of the constraint, \( Cond \) is the application condition of the constraint, and \( Pred \) its predicate or succeedent. The logical meaning of the constraint is that for each instance \( self \) of \( S_i \) which exists in the source model, if \( self \) satisfies \( Cond \), then \( Pred \) is true for \( self \). The operational meaning of the constraint is that for each instance \( self \) of \( S_i \) which exists in the source model, if \( self \) satisfies \( Cond \), then \( Pred \) is made true for \( self \), using the activity \( \text{stat}_L(C)(P) \).

Typically, \( Pred \) will specify the existence of instances \( t \) of target entity types \( T_j \) via formulae \( T_j \rightarrow \text{exists}(t|P) \) for concrete entity type \( T_j \), where \( t \)'s feature values are specified in \( P \). For example, the simple Families to Persons [1] forward transformation \( \text{family2person} \) (Figure 1 shows the basic metamodels of this transformation) could have two rules:

**Family::**

\[m : \text{mother} \rightarrow \text{union}(\text{daughters}) =>\]

\[\text{Female} \rightarrow \text{exists}(p | \]

\[p.\text{name} = \text{name} + ", " + m.\text{name} )\]

**Family::**

\[m : \text{father} \rightarrow \text{union}(\text{sons}) =>\]

\[\text{Male} \rightarrow \text{exists}(p | \]

\[p.\text{name} = \text{name} + ", " + m.\text{name} )\]

These rules \( Post_{\text{family2person}} \) express that for each Family instance, and for each FamilyMember \( m \) which is either a father or son of the family, there should exist a corresponding Male instance with a name formed as the concatenation of
FIGURE 1. Family to Person metamodels.

the family and family member names. Likewise, mothers and daughters are mapped to `Female` instances. The invariants \( \text{Inv}_{\text{family2person}} \) can be derived syntactically from these rules by inverting the quantifiers (Table 5, case 2):

Female::
\[
\text{Family}\rightarrow\text{exists}(\text{fx} | \text{FamilyMember}\rightarrow\text{exists}(\text{m} | \text{m} : \text{fx}.\text{father}\rightarrow\text{union}(\text{fx}.\text{sons}) \& \text{name} = \text{fx}.\text{name} + "", " + \text{m}.\text{name} ) )
\]

Male::
\[
\text{Family}\rightarrow\text{exists}(\text{fx} | \text{FamilyMember}\rightarrow\text{exists}(\text{m} | \text{m} : \text{fx}.\text{mother}\rightarrow\text{union}(\text{fx}.\text{daughters}) \& \text{name} = \text{fx}.\text{name} + "", " + \text{m}.\text{name} ) )
\]

These invariants are preserved by the `family2person` rules: if a new `Female` instance is created by the first rule, then it necessarily satisfies the first invariant because of the context entity type and application conditions of the rule. The new instance does not affect the truth of the second invariant, which only restricts `Male` instances. Likewise, the second rule preserves both invariants.

The importance of transformation invariants in verification is that they can be used to prove transformation contracts: conditions \( P \) and \( Q \) such that, if transformation \( \tau \) is executed on models satisfying \( P \), it is ensured to terminate in a state with the models satisfying \( Q \) [7], [19]. If \( I \) is a condition such that \( P \) implies \( I \), with \( I \) preserved by each rule application of \( \tau \), and such that \( I \) together with the fact that no rule is enabled implies \( Q \), then the correctness part of the contract is established.

The key idea of our first bx approach is that change-propagation from the source model to the target can be achieved by extending a forward transformation \( \tau \) by additional constraints/rules derived automatically from \( \text{Inv}_\tau \). In addition, a reverse transformation \( \tau^- \) can also be derived automatically from \( \text{Inv}_\tau \). The reverse transformation is based on interpreting the invariants as rules in the inverse direction (notice that the form of the invariants in the `family2person` example is very similar to that of use case postconditions).

A rule predicate \( P \) such that \( \text{t.name} = \text{s.name} + "\*" \) is converted to an explicit inverse form \( P^- \) such as \( \text{s.name} = \text{t.name.front} \) in order to provide a computationally explicit version of the invariants. This approach therefore achieves the aims of bx in enabling a single transformation specification to be used for both forward and reverse transformations.

B. TRANSFORMATION IMPLEMENTATION

Designs for model transformations and for general applications are defined in UML-RSDS using class diagrams, with UML activities expressing the explicit algorithms of operations and use cases.

The following concrete syntax is used for a subset of UML structured activities:

\[
<\text{statement}> ::= <\text{loop}> | <\text{creation}> | <\text{conditional}> | <\text{sequence}> | <\text{choice}> | <\text{basic}>
\]

\[
<\text{loop}> ::= "\text{while}" <\text{expression}> | "\text{do}" <\text{statement}> | "\text{for}" <\text{expression}> | "\text{do}" <\text{statement}>
\]

\[
<\text{conditional}> ::= "\text{if}" <\text{expression}> | "\text{then}" <\text{statement}> | "\text{else}" <\text{basic}>
\]

\[
<\text{sequence}> ::= <\text{statement}> | "\text{;}" <\text{statement}>
\]

\[
<\text{choice}> ::= <\text{statement}> | "\text{[}" <\text{statement}> | "\text{;}" <\text{identifier}>
\]

\[
<\text{basic}> ::= <\text{expression}> | "\text{=}" <\text{expression}> | "\text{skip}" | "\text{execute}" | "\text{return}"
\]

For use cases, their procedural implementation as an activity is derived as a composition of the activities \( \text{stat}_\text{LC}(Cn) \) for each postcondition constraint \( Cn.\;\text{stat}_\text{LC}(Cn) \) is constructed.
so that it establishes \(C_n: [stat_{LC}(C_n)]C_n\), where \([act]P\) is the weakest-precondition of predicate \(P\) with respect to statement/activity \(act\). As a predicate transformer, \(stat_{LC}(C_n)\) is chosen as a minimal (least-refined) transformer which establishes \(C_n\). In terms of the refinement calculus [50]:

\[
[C]P \implies (stat_{LC}(P) \subseteq C)
\]

for statements \(C\) which do not update more variables than \(stat_{LC}(P)\). Intuitively, this means that \(stat_{LC}(P)\) performs the minimal necessary data modifications needed to establish \(P\), operating on the write frame \(wr(P)\) of \(P\). A definition of \(stat_{LC}\) is given in Section III-C. There may not be a unique least-refined statement that establishes a given predicate, in such cases a suitable minimally-refined statement is selected instead.

Data-dependency relations between the \(Post\) constraints are important in ensuring the correctness of both unidirectional and bidirectional UML-RSDS transformations. Let \(rd(C_n)\) denote the read-frame of a constraint \(C_n\), the set of entity type names and data features that it reads, and \(wr(C_n)\) denote its write frame. These are defined in [36]. The definition of \(wr(P)\) for a predicate \(P\) interpreted operationally as \(stat_{LC}(P)\), includes features and entity types affected by implicit updates which may arise due to metamodel constraints involving parts of the metamodel explicitly updated by \(stat_{LC}(P)\). There are five main cases of such implicit updates:

1. A bi-directional association \(A m_1 \xrightarrow{\text{ar}} m_2 B\) between entity types \(A\) and \(B\), for any multiplicities \(m_1, m_2\); changes to \(ar\) will also produce updates to \(br\) and vice-versa.
2. An association \(A m_1 \xrightarrow{\text{br}} B\): deletion of \(B\) instances will remove these from \(br\).
3. An association \(A 1 \xrightarrow{\text{br}} B\) with 1-multiplicity end at entity type \(A\): deletion of \(A\) instances \(ax\) will result in deletion of all \(B\) instances linked to \(ax\) via \(br\). For example, in Figure 1, deletion of a Family instance leads to deletion of its linked FamilyMembers.
4. A composition aggregation: deletion of the composite will result in deletion of its linked parts.
5. Creation or deletion of instances of a subclass \(B\) of a class \(A\) will also result in addition/ removal of these instances from the extent of \(A\).

In Figure 1, deletion of Family instances will also affect FamilyRegister :: families, even though this feature is not explicitly modified by the transformation. Likewise for Person and PersonRegister :: persons. We say that there is deletion propagation from entity type \(E_i\) to entity type \(E_j\) if there is a non-empty chain of dependencies of kinds 3, 4 and 5 above between \(E_i\) and \(E_j\) such that deletion of an \(E_i\) instance may cause deletion of an \(E_j\) instance.

A dependency ordering \(C_n < C_m\) is defined between constraints by \(wr(C_m)\cap rd(C_n) \neq \{\}\) “\(C_m\) depends on \(C_n\)”. A use case with postconditions \(C_1, \ldots, C_n\) should satisfy the syntactic non-interference conditions:

1. If \(C_i < C_j\), with \(i \neq j\), then \(i < j\).
2. If \(i \neq j\) then \(wr(C_i) \cap wr(C_j) = \{\}\).

Together, these conditions ensure that the activities \(stat_{LC}(C_i)\) of subsequent constraints \(C_j\) cannot invalidate earlier constraints \(C_i\), for \(i < j\).

A constraint is termed a type 1 constraint if its write and read frames are disjoint. Such constraints usually have an implementation as a bounded loop, as opposed to a fixed-point/unbounded iteration. Both of the rules for Family to Person above are of type 1.

Entity types may have identity attributes of String type. These uniquely identify instances of the entity: two different instances must have different values of the attribute. The first such attribute is used as a primary key for the entity type.

If entity type \(E\) has a primary key \(att : String\), then \(E[v]\) denotes the instance \(x\) of \(E\) with key value \(x.att\) equal to \(v\), or the collection of \(E\) instances with key value in \(v\), if \(v\) is a collection.

**C. Definition of Activities for Predicates**

Table 1 gives the definition of \(stat_{LC}\) for simple predicates.

**TABLE 1. Definition of \(stat_{LC}\) for simple predicates.**

<table>
<thead>
<tr>
<th>Predicate (P)</th>
<th>(stat_{LC}(P))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x = e)</td>
<td>(x := e)</td>
</tr>
<tr>
<td>Assignable basic expression (x)</td>
<td>(x := x\rightarrow including(e))</td>
</tr>
<tr>
<td>(e : x)</td>
<td>(x := x\rightarrow including(e))</td>
</tr>
<tr>
<td>Assignable basic expression (x), Set-valued</td>
<td>(if x\rightarrow includes(x) ) then skip else (x := x\rightarrow append(x))</td>
</tr>
<tr>
<td>(e / : x)</td>
<td>(x := x\rightarrow excluding(x))</td>
</tr>
<tr>
<td>Assignable basic expression (x)</td>
<td>for (x : e) do (stat_{LC}(x : f))</td>
</tr>
<tr>
<td>(e \ll f)</td>
<td>(f := f \rightarrow e)</td>
</tr>
<tr>
<td>Assignable basic expression (f)</td>
<td></td>
</tr>
</tbody>
</table>

An assignable basic expression \(x\) is either a variable, a non-frozen feature \(f\) or a basic expression \(obj.f\) or \(obj.f[i]\) whose modified feature \(f\) is non-frozen. \(stat_{LC}(x \rightarrow includes(e))\) is the same as \(stat_{LC}(e : x)\), and likewise \(x \rightarrow excludes(e)\) is considered the same as \(e / x \rightarrow includesAll\) if the same as \(e \ll x\).

\(stat_{LC}(true)\) is skip. For \(x \rightarrow is Deleted\), where \(x\) is an instance of class \(E\), \(stat_{LC}\) removes \(x\) from the set of instances of \(E\), and performs all the deletion propagation actions from \(E\) listed in the preceding subsection.

\(stat_{LC}(P \& Q)\) is \(stat_{LC}(P)\); \(stat_{LC}(Q)\) under the assumption that \(P \implies [stat_{LC}(Q)P]\).
\[ \text{stat}_{LC}(\text{if } C \text{ then } P \text{ else } Q \text{ endif}) \text{ is} \]
\[ \text{if } C \]
\[ \text{then } \text{stat}_{LC}(P) \]
\[ \text{else } \text{stat}_{LC}(Q) \]

under the assumptions \( C \implies [\text{stat}_{LC}(P)]C \) and \( \neg C \implies [\text{stat}_{LC}(Q)]\neg C \).

For \( s \rightarrow \text{forall}(x|P) \), \( \text{stat}_{LC} \) is defined as for \( x : s \) do \( \text{stat}_{LC}(P) \), when \( rd(s) \cap wr(P) = \{ \} \) and \( rd(P) \cap wr(P) = \{ \} \).

For existential quantifiers \( E \rightarrow \text{exists}(e|P_1 \& \ldots \& P_n) \), their \( \text{stat}_{LC} \) effect only creates a new \( e : E \) instance in cases where there is no existing \( e : E \) that satisfies \( P \) partially or completely. In the case of partial satisfaction the updates only for the unsatisfied conjuncts are carried out.

If \( E \) has a primary key \( pk \) and a conjunct \( P_i \) is of the form \( e.pk = \text{value} \), then \( \text{stat}_{LC}(E \rightarrow \text{exists}(e|P_1 \& \ldots \& P_n)) \) is
\[ \text{var } e : E; \]
\[ e := \text{createByPKE}(\text{value}); \]
\[ \text{stat}_{LC}(\text{Pred}) \]

Where \( \text{Pred} \) is \( P_1 \& \ldots \& P_n \) with \( P_i \) omitted. \( \text{createByPKE}(\text{val}) \) looks up \( E[\text{val}] \), if this is non-null it returns the instance, otherwise it creates a new \( E \) instance with primary key value \( \text{val} \), and returns this instance.

Otherwise, \( \text{stat}_{LC}(E \rightarrow \text{exists}(e|P_1 \& \ldots \& P_n)) \) has the form:
\[ \text{var } e : E; \]
\[ \text{var eset : Set}(E); \]
\[ \text{eset} := E\.\text{allInstances}(); \]
\[ \text{if eset}\rightarrow\text{isEmpty}() \]
\[ \text{then} \]
\[ e := \text{createE}(); \]
\[ \text{stat}_{LC}(P_1 \& \ldots \& P_n) \]
\[ \text{else} \]
\[ (e := \text{eset}\rightarrow\text{any}()); \]
\[ \text{eset} := \text{eset}\rightarrow\text{select}(P_1); \]
\[ \text{if eset}\rightarrow\text{isEmpty}() \]
\[ \text{then} \]
\[ e := \text{createE}(); \]
\[ \text{stat}_{LC}(P_1 \& \ldots \& P_n) \]
\[ \text{else} \]
\[ (e := \text{eset}\rightarrow\text{any}()); \]
\[ \text{eset} := \text{eset}\rightarrow\text{select}(P_2); \]
\[ \text{if eset}\rightarrow\text{isEmpty}() \]
\[ \text{then} \]
\[ \text{stat}_{LC}(P_2 \& \ldots \& P_n) \]
\[ \text{else} \]
\[ \ldots \text{case for } 3 \ldots ) \]

The general case for \( k \geq 2, k < n \) is
\[ e := \text{eset}\rightarrow\text{any}(); \]
\[ \text{eset} := \text{eset}\rightarrow\text{select}(P_k); \]
\[ \text{if eset}\rightarrow\text{isEmpty}() \]
\[ \text{then} \]
\[ \text{stat}_{LC}(P_k \& \ldots \& P_n) \]
\[ \text{else} \]
\[ \ldots \text{case for } k + 1 \ldots \]

For \( k = n \), if \( P_k \) is an assignment \( \text{result} = e \), then the code of the case is simply \( e := \text{eset}\rightarrow\text{any}(); \text{result} := e \). Otherwise it is
\[ e := \text{eset}\rightarrow\text{any}(); \]
\[ \text{eset} := \text{eset}\rightarrow\text{select}(P_n); \]
\[ \text{if eset}\rightarrow\text{isEmpty}() \]
\[ \text{then} \]
\[ \text{stat}_{LC}(P_n) \]
\[ \text{else} \]
\[ e := \text{eset}\rightarrow\text{any}() \]

By reusing \( e : E \) instances where possible, the redundant creation of instances is avoided, however this also introduces the possibility of conflicts where one target instance is required to have conflicting attribute values to satisfy a constraint wrt 2 source instances.

**IV. PATTERNS FOR MODEL TRANSFORMATIONS**

In [38] we identified 29 design patterns which have been used for model transformation specification and design. According to [43], some of the most widely-used MT patterns in practice are Entity Splitting, Structure Preservation, Map Objects before Links and Entity Merging. Our approach to bx specification using UML-RSDS applies to transformations \( \tau \) constructed using these patterns and others, and enables the automatic derivation of inverse transformations \( \tau^{-1} \) which also use the patterns, according to Table 2.

**TABLE 2. Correspondence of MT patterns.**

<table>
<thead>
<tr>
<th>Pattern in ( \tau )</th>
<th>Pattern in ( \tau^{-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity Splitting</td>
<td>Entity Merging</td>
</tr>
<tr>
<td>Entity Merging</td>
<td>Entity Splitting</td>
</tr>
<tr>
<td>Map Objects before Links</td>
<td>Map Objects before Links</td>
</tr>
<tr>
<td>Structure Preservation</td>
<td>Structure Preservation</td>
</tr>
<tr>
<td>Phased Construction</td>
<td>Phased Construction</td>
</tr>
<tr>
<td>Flattening</td>
<td>Unflattening</td>
</tr>
<tr>
<td>Unflattening</td>
<td>Flattening</td>
</tr>
<tr>
<td>Inverse Recursive Descent</td>
<td>Inverse Recursive Descent</td>
</tr>
</tbody>
</table>

Here we give a more precise characterization of the patterns, and explain their specialization for use in bx. In each case in the following patterns, the predicates \( TCond, P \) contain no quantifiers or calls to update operations, and
have an operational interpretation \( \text{stat}_{LC} \). They can contain calls of purely functional operations that do not create instances or modify data.

- **Structure Preservation**: a rule in which one source entity type \( S_i \) is mapped to one target entity type \( T_j \), with no dependency upon any other target entity types. Rules \( C_i \) have the form

\[
S_i :\quad S\text{Cond}_i(\text{self}) \implies T_j \exists \text{exists}(t | T\text{Cond}_i(t) \land P_i(\text{self}, t))
\]

where \( S\text{Cond}_i \) refers only to source model entity types and features. In the case of a bx, identity attributes \( idS_i \) and \( idT_j \) of the entity types can be used to correlate instances of \( S_i \) and \( T_j \).

- **Phased Construction**: as for Structure Preservation, but with previously-created instances of target entity types \( T\text{Ref} \) derived from \( S\text{Ref} \) instances being looked up in the succedent of a rule creating an instance of target entity \( T_j \), in order to define the value of a \( T\text{Ref}-\text{typed} \) role \( rr \) of the \( T_j \) instance, as in Figure 2. \( P_j \) in this case contains some equation \( t.rr = T\text{Ref}[r.idS\text{Ref}] \). For example:

\[
S_i :\quad S\text{Cond}_i(\text{self}) \implies T_j \exists \text{exists}(t | T\text{Cond}_i(t) \land P_j(\text{self}, t) \land T\text{Ref}[r.idS\text{Ref}])
\]

This can be used for single-valued \( r \), \( rr \) as well as for collection-valued association ends.

- **Entity Splitting**: this pattern involves one source entity type mapping to two or more different target entity types. There are two variants of this pattern: (i) **horizontal**: two or more separate rules map from \( S_i \), with disjoint \( S\text{Cond}_i \) application conditions, to distinct target entity types \( T_k \), \( T_i \); (ii) **vertical**: a single rule maps one \( S_i \) instance to multiple linked instances of different \( T_k \) \( T_i \). The second case is common in refinement transformations. Constraints, eg., \( C_i \), will have the form

\[
S_i :\quad S\text{Cond}_i(\text{self}) \implies T_k \exists \text{exists}(t | T\text{Cond}_i(t) \land T\text{Cond}_2(t_1, t_2) \land P(\text{self}, t_1, t_2))
\]

Some explicit means of recording the semantic link between \( t_1 \) and \( t_2 \) is needed to support bx definition for such transformations, this link data can be based on identity attributes. This should be the only constraint in the transformation that creates \( T_i \) instances. Other constraints can create \( T_k \) instances if their \( T\text{Cond} \) conditions are disjoint from \( T\text{Cond}_1 \).

- **Entity Merging**: As with Entity Splitting, there are two variants, which are inverses to the corresponding versions of Entity Splitting. In **horizontal** Entity Merging, two or more source entity types \( S_i, S_j \) are mapped to the same target entity type \( T_k \) by separate rules:

\[
S_i :\quad S\text{Cond}_i(\text{self}) \implies T_k \exists \text{exists}(t | T\text{Cond}_i(t) \land P_i(\text{self}, t))
\]

and

\[
S_j :\quad S\text{Cond}_j(\text{self}) \implies T_k \exists \text{exists}(t | T\text{Cond}_j(t) \land P_j(\text{self}, t))
\]

For bx, sufficient information must be inserted into the \( T_k \) instances to enable the source entity of the instance \( S_i \) or \( S_j \) to be identified. Usually \( T\text{Cond}_i \) and \( T\text{Cond}_j \) would be logically disjoint for this reason:

\[
T\text{Cond}_i \equiv \text{not}(T\text{Cond}_j)
\]

In **vertical** Entity Merging, two or more source instances are used to construct a single target instance, in a single rule:

\[
S_i :\quad S\text{Cond}_1(\text{self}) \land s_2 : S_2 \land S\text{Cond}_2(\text{self}, s_2) \implies T \rightarrow \exists \text{exists}(t | T\text{Cond}_1(t) \land P(\text{self}, s_2, t))
\]

This rule iterates over pairs \( s_1 : S_1, s_2 : S_2 \), potentially creating a \( t : T \) instance for each pair that satisfies the antecedent conditions.

- **Map Objects before Links**: This pattern, also known as Entities before Relations [49], separates the construction of target instances from the linking of target instances based on source links (Figure 2).

![FIGURE 2. Map Objects before Links pattern.](image)

Entity mappings are as for any of the above patterns, whilst the linking constraints/rules have the form:

\[
T_j[idS_i].rr = T\text{Ref}[r.idS\text{Ref}]
\]

These define target model association ends \( rr \) from source model association ends \( r \), looking-up target
model elements $T_1[idS_1]$ and $TRef[v.idSRef]$ which have already been created by preceding entity mapping rule(s).

The Families to Persons transformation illustrates Entity Splitting (variant (i)). In the forward direction, FamilyMember instances are mapped to Male instances if they occur in the father or sons collections of some family, and to Female instances if they occur in the mother or daughters collections. This is an example of Entity Splitting variant (i): an instance of one source entity is mapped to one instance of possibly different alternative target entities under disjoint assumptions.

We have also identified a number of specialized patterns for bx, based on inspection of published examples of bx [43]:

- **Auxiliary Correspondence Model**: using auxiliary entity types or features, maintain an explicit correspondence between source model and target model elements to facilitate change-propagation in source to target or target to source directions (Figure 3). This pattern is a built-in mechanism of the TGG language and has been used in numerous transformation cases [3], [16], [17], [21], [22], [24], [25], [27], [29], [32], [41], [55], [58], [67]. In QVT-R there are implicit trace entities for each relation [51].

- **Cleanup before Construct**: to enable change-propagation for $R^−$, define additional rules that remove superfluous target model elements which are not in the bx relation $R$ with any source elements, before constructing target elements related to source elements, and similarly for $R^−$ (Figure 4). This pattern involves the introduction of additional rules, prior to the main rules of the transformation, to remove target elements which fail to correspond to source elements. Alternatively, cleanup actions can be performed after the main transformation, as with QVT-R delete actions [51].

- **Unique Instantiation for bx**: With this pattern, the $E→exists(e|P)$ quantifier in rule succedents, for concrete entity type $E$, is interpreted as “create a new instance $e$ of $E$ and establish $P$ for $e$, unless there already exists an $e : E$ satisfying $P$”. This corresponds to the ‘check before enforce’ semantics for QVT-R [59].

As described in Section III-C, in the case that $E$ has a primary key $eld$, $E→exists(e|eld = v & P)$ is interpreted as the sequence of steps: (i) lookup $E[v]$, if this exists then assign it to $e$, otherwise create a new $E$ instance with $eld = v$ and assign this to $e$; (ii) try to establish $P$ for $e$, taking no action if $P$ already holds.

This avoids the deletion and recreation of elements $t$ in one model which correspond by key to an element $s$ in the other model, but which have divergent data. Instead the data of $t$ is modified to enforce the transformation relation.

If $E$ does not have a primary key, then $statE_c$ ($E→exists(e|P)$) selects an $e : E$ for which a maximal initial part $P_1$ of $P$ holds and tries to establish the remainder of $P$ for this $e$, or it creates a new $e$ and establishes $P$ for it, if no element of $E$ satisfies any initial part $P_1$ of $P$.

- **Lens**: the source to target mapping is independent of the target model data, with the target model $t : TL$ computed as a view or abstraction $get(s)$ containing partial information from the source model $s : SL$ [15], [6]. A dual map $put : SL \times TL → SL$ propagates view changes to the source, such that $put(s, get(s)) = s$ and $get(put(s, t)) = t$.

In our formalism we define view-update functions $put_E$ for views $get_E$ in order to reverse fine-grained data relations $P(s, t)$ such as $t.g = get_E(s.f)$ between source model data $s.f$ and target data $t.g$, for a wide range of expression forms $get_E$, as in [34] and [61]. The inverse $P^−(s, t)$ of $P(s, t)$ is defined as $s.f = put_E(s.f @pre, t.g)$ in such a case.

Our approach to implement Auxiliary Correspondence Model is to introduce String-valued identity attributes (primary keys) for source and target entities. These attributes are used to record which source and target instances correspond: target instance $t$ corresponds to source $s$ if they have the same primary key values. For example, the person to family case would be extended by this pattern as shown in Figure 5.

In addition, we introduce here two new bx patterns, which we have used in a large-scale code generation transformation (Section VII):
Inverse Recursive Descent: if a forward transformation \( \tau \rightarrow \) is defined using structural recursion on the source language \( SL \), the reverse transformation \( \tau \leftarrow \) may be derived from \( \tau \rightarrow \) as a dual structural recursion on \( TL \) (Figure 6). If source entity \( E \) has subelements \( esub : Collection(ESub) \), and the corresponding target entity \( F \) has corresponding subelements \( fsub : Collection(FSub) \), then a forward rule

\[
E ::
true \rightarrow self.op1()
\]

based on operations \( op1 \) and \( op \) of \( E \) of the form

\[
E ::
op1() : FSup
\]

\[
pre : true
\]

\[
post : result = op(esub.op1())
\]

\[
E ::
op(fs : Collection(FSub)) : F
\]

pre : \( fs = FSub[esub.eld] \)

post :

\[ECond \rightarrow F \rightarrow \exists f \left( f.Fld = fId \& FCond(f) \& f.fsub = fs \& P(self,f) \& result = f \right)\]

has the transformation invariant

\[
F ::
FCond \rightarrow E \rightarrow \exists e \left( e.eld = fId \& ECond(e) \& e.esub = ESub[fsub.fld] \& P(e,self) \right)
\]

provided that \( F \) is only created/modified by \( op \).

The inverse operations for the reverse transformation are:

\[
F ::
op1\sim() : FSup
\]

\[
pre : true
\]

\[
pot : result = \sim\left( fsub.op1\sim() \right)
\]

\[
E ::
op(es : Collection(ESub)) : E
\]

\[
pre : es = ESub[fsub.fld]
\]

\[
pot : FCond \rightarrow E \rightarrow \exists e \left( e.eld = fId \& ECond(e) \& e.esub = es \& P\sim(e,self) \& result = e \right)
\]
The enhanced rules of family2person using the Auxiliary Correspondence Model and Flattening/Unflattening patterns are then:

\[
\text{Family::}
\begin{align*}
  m & : \text{mother->union(daughters)} => \\
  \text{Female->exists} & (p | \\
  p.\text{personId} = m.\text{memberId} & \\
  p.\text{familyId} = id & \\
  p.\text{name} = \text{name} + "," + m.\text{name})
\end{align*}
\]

There are corresponding derived invariants. Other forms of flattening discard certain model elements (eg., if only leaf classes in a class diagram are mapped to relational tables) and associations between elements (e.g., state composition links in the state machine flattening example of [46]). The reverse transformation then needs to reconstruct the model elements or structure.

We can extend the patterns of Table 2 to include update operation calls within succedents, provided the operations are invertible in the manner of the inverse recursive descent pattern above. An operation defined as

\[
\begin{align*}
  S_i :: \\
  op() : T_j \\
  \text{post} : \\
  S\text{Cond}(self) & \implies \\
  T_j & \implies \text{exists}(t | T\text{Cond}(t) & P_j(\text{self}, t) & \text{result} = t) \\
\end{align*}
\]

can be inverted to:

\[
\begin{align*}
  T_j :: \\
  op^-() : S_i \\
  \text{post} : \\
  T\text{Cond}(self) & \implies \\
  S_i & \implies \text{exists}(s | S\text{Cond}(s) & P_j^-(s, \text{self}) & \text{result} = s)
\end{align*}
\]

Applications \(y.rr = x.r.op()\) in a constraint succedent invert to \(x.r = y.rr.op^-()\).

If \(T_j\) elements are only created/modified by \(op\), then the transformation satisfies the invariant:

\[
\begin{align*}
  T_j :: \\
  T\text{Cond}(self) & \implies \\
  S_i & \implies \text{exists}(s | S\text{Cond}(s) & P_j(s, \text{self}))
\end{align*}
\]

V. APPLICATION OF PATTERNS TO CONSTRUCT BX

A unidirectional transformation \(t_0\) defined using the patterns of Table 2, and consisting of type 1 constraints satisfying syntactic non-interference, can be extended by the use of the Auxiliary Correspondence Model and Cleanup before Construct patterns to a transformation \(\tau^{-}\) which is the forward part of a bx. The bx relation and reverse transformation \(\tau^{+}\) can also be derived.

An alternative condition to syntactic non-interference is semantic non-interference: a use case with postconditions \(C_1\) to \(C_n\) satisfies this property if for \(i < j\):

\[
C_i \implies \text{stat}_{LC}(C_j)C_i
\]
TABLE 3. Inverses of predicates and expressions.

<table>
<thead>
<tr>
<th>$P(s,t)$</th>
<th>$P^\rightarrow(s,t)$</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t.g = s.f$</td>
<td>$s.f = t.g$</td>
<td>Assignable features $f, g$</td>
</tr>
<tr>
<td>$t.g = s.f \cdot f$</td>
<td>$s.f = t.g \cdot f$</td>
<td>Invertible function/composition $f, g$, $E.g., (cosh(\log(x)) = exp(acosh)$</td>
</tr>
<tr>
<td>$t.x = \begin{cases} C(s.x) &amp; \text{then } f(s.x) \ \text{else } g(s.x) \end{cases}$</td>
<td>$s.x = \begin{cases} C'(t.y) &amp; \text{then } f'(t.y) \ \text{else } g'(t.y) \end{cases}$</td>
<td>$f$ on $C$, $g$ on $not(C)$ are invertible with disjoint ranges. $C'(z)$ expresses that $z$ is in the range of $f$ on $C$.</td>
</tr>
<tr>
<td>$t.g = 1/s.f \cdot f$</td>
<td>$s.f = (1/t.g) \cdot f^{-1}$</td>
<td>Invertible double-valued $f$</td>
</tr>
<tr>
<td>$t.g = s.f \cdot pow(x)$</td>
<td>$s.f = t.g \cdot pow(1.0/x)$</td>
<td>$x$ non-zero, independent of $s.f$</td>
</tr>
<tr>
<td>$t.g = a \cdot fpow(s.f)$</td>
<td>$s.f = t.g \cdot flog(a \cdot \log(s.f))$</td>
<td>$a &gt; 0$, constant</td>
</tr>
<tr>
<td>$t.f = \log()$</td>
<td>$s.f = t.g \cdot log()$</td>
<td>$t.g &gt; 0$</td>
</tr>
<tr>
<td>$t.f = t.g \cdot asin()$</td>
<td>$s.f = t.g \cdot cos()$</td>
<td>$-1 \leq t.g \leq 1$</td>
</tr>
<tr>
<td>$t.f = t.g \cdot acos()$</td>
<td>$s.f = t.g \cdot tan()$</td>
<td>$-1 \leq t.g \leq 1$</td>
</tr>
<tr>
<td>$t.b2 \cdot not(s.b1)$</td>
<td>$s.b1 = not(t.b2)$</td>
<td>Boolean attributes $b1, b2$</td>
</tr>
<tr>
<td>$t.g = K \cdot s.f + L$</td>
<td>$s.f = (t.g - L)/K$</td>
<td>$f, g$ numeric attributes</td>
</tr>
<tr>
<td>$t.g = s.f + S K$</td>
<td>$s.f = t.g \cdot \text{subrange}(1, t.g \cdot \text{size} - S \cdot K)$</td>
<td>String constant $S K$</td>
</tr>
<tr>
<td>$t.g = S + s.f$</td>
<td>$s.f = t.g \cdot \text{subrange}(S \cdot K + 1, t.g \cdot \text{size})$</td>
<td>String constant $S K$</td>
</tr>
<tr>
<td>$t.f = s.f \cdot reverse()$</td>
<td>$s.f = t.g \cdot reverse()$</td>
<td>String valued $f$</td>
</tr>
<tr>
<td>$t.f = \text{StringLib.rotate}(s.f, n)$</td>
<td>$s.f = \text{StringLib.rotate}(t.g, n)$</td>
<td>String valued $f$</td>
</tr>
<tr>
<td>$R(s,t)$ &amp; $Q(s,t)$</td>
<td>$R^{-1}(s,t)$ &amp; $Q^{-1}(s,t)$</td>
<td>$E$ is the entity element type of $s.f$</td>
</tr>
</tbody>
</table>

Syntactic non-interference implies semantic non-interference, but not conversely.

The bx derivation process consists of the following steps to enhance $t_0$:

1) Introduce identity attributes in source and target entity types where necessary to support the Auxiliary Correspondence Model and Flattening/Unflattening patterns, and extend the postconditions to use these attributes, forming extended postconditions $Post_\tau$ of an enhanced transformation $\tau$ of $t_0$ (cf., the step from Figure 1 to Figure 5).

2) Compute the invariant $Inv_\tau$ of this extended transformation.

3) The conjunction $Post_\tau$ & $Inv_\tau$ is the bx relation $R_\tau$ of the extended transformation (and the logical relation of the ‘transformation model’ of $\tau$ in the sense of [9]).

4) Further extend $\tau$ to a forward transformation $\tau^{-1}$ by applying Cleanup before Construct to add rules (applied before the main rules of $\tau$) which remove target instances that fail to satisfy $Inv_\tau$ with respect to corresponding source instances.

5) Compute an explicit form $Post_\tau^{-1}$ of $Inv_\tau$, to form the basis $\tau^{-1}$ of the reverse transformation $\tau^\rightarrow$. This step requires that the computations $t.g = getE(s.f)$ of target data $t.g$ in terms of source data $s.f$ used within the postcondition predicates of $\tau$ are invertible using view update functions $putE$ in the sense of a lens.

For transformation postconditions conforming to the patterns of Table 2 and consisting only of type 1 constraints, using invertible computations of target data based on source data which are known as valid get/put pairs, $Inv_\tau$ and $Post_\tau^{-1}$ can be derived mechanically from $Post_\tau$ using the definition of invariant predicates and inverse predicates $P^{-1}$ given in Tables 3, 4, 5, 6 and in other catalogued cases. Steps 2) to 5) can therefore be automated for transformations satisfying these restrictions. Note that in the final case of Table 3 we permit a quantification of the form $\forall x \exists y$ within the $P^{-1}$ predicate of a constraint, thus extending slightly the range of constraints to which the bidirectionalization procedure can be applied.

Given a type 1 structure preservation rule $C_i$ of the form

$S_i ::$

$\text{SCond}(self) \implies T_j \rightarrow \exists \forall x \mid P(x,y) \land y : t.g)$

the corresponding invariant constraint $Inv_i$ is:

$T_j ::$

$\text{TCond}(self) \implies S_i \rightarrow \exists \forall x \mid P^{-1}(x,y) \land x : s.f)$

The invariant derived from the rule expresses that each target model instance of $T_j$ corresponds to some source model instance of $S_i$ via that rule.

Tables 3 and 4 show some examples of inverses $P^{-1}$ of expressions and predicates $P$. The computation of these inverses is implemented in the UML-RSDS tools as a higher-order transformation (the reverse option for use cases). Interception on sequences is treated as select:
TABLE 4. Inverse of predicates on associations.

<table>
<thead>
<tr>
<th>$P(s,t)$</th>
<th>$P^{-1}(s,t)$</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{rr} = s.r \rightarrow \text{reverse()}$</td>
<td>$s.r = t_{rr} \rightarrow \text{reverse()}$</td>
<td>$r, rr$ ordered association ends</td>
</tr>
<tr>
<td>$t_{rr} = s.r \rightarrow \text{first()}$</td>
<td>$s.r \rightarrow \text{first()} = t_{rr}$</td>
<td>$r$ ordered association end</td>
</tr>
<tr>
<td>$t_{rr} = s.r \rightarrow \text{last()}$</td>
<td>$s.r \rightarrow \text{last()} = t_{rr}$</td>
<td>$r$ ordered association end</td>
</tr>
<tr>
<td>$t_{rr} = s.r \rightarrow \text{including}(s.p)$</td>
<td>$s.r = t_{rr} \rightarrow \text{front()} &amp; s.p = t_{rr} \rightarrow \text{last()}$</td>
<td>$p_1$-multiplicity end</td>
</tr>
<tr>
<td>$t_{rr} = \text{Sequence}{s.p1, s.p2}$</td>
<td>$s.p1 = t_{rr} \rightarrow \text{at}(1) &amp; s.p2 = t_{rr} \rightarrow \text{at}(2)$</td>
<td>$rr$ ordered association end</td>
</tr>
<tr>
<td>$t_{rr} = s.r \rightarrow \text{select}(P)$</td>
<td>$s.r = s.r@\text{pre} \rightarrow \text{reject}(P) \cup t_{rr} \rightarrow \text{select}(P)$</td>
<td>$r, rr$ set-valued</td>
</tr>
<tr>
<td>$t_{rr} = s.r \rightarrow \text{sort()}$</td>
<td>$s.r = t_{rr} \rightarrow \text{hasSet()}$</td>
<td>$r$ set-valued, $rr$ sequence-valued</td>
</tr>
<tr>
<td>$t_{rr} = s.r \rightarrow \text{hasSequence()}$</td>
<td>$s.r = (s.r@\text{pre} \land t_{rr}) \Rightarrow (t_{rr} - s.r@\text{pre})$</td>
<td>$r, rr$ sequence-valued</td>
</tr>
</tbody>
</table>

TABLE 5. Inverse of constraints.

<table>
<thead>
<tr>
<th>$Cn$</th>
<th>$Cn^{-}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{i}::$</td>
<td>$T_{j}::$</td>
</tr>
<tr>
<td>$SCond(self) \Rightarrow T_{j} \rightarrow \text{exists}(t \mid \text{TCond}(t) &amp; P(self,t))$</td>
<td>$TCond(self) \Rightarrow S_{i} \rightarrow \text{exists}(s \mid SCond(s) &amp; P^{-1}(s, self))$</td>
</tr>
<tr>
<td>$S_{i}::$</td>
<td>$T_{j}::$</td>
</tr>
<tr>
<td>$v : F &amp; SCond(self, v) \Rightarrow T_{j} \rightarrow \text{exists}(t \mid \text{TCond}(t) &amp; P(self,v,t))$</td>
<td>$TCond(self) \Rightarrow S_{i} \rightarrow \text{exists}(s \mid F \rightarrow \text{exists}(v \mid SCond(s,v) &amp; P^{-1}(s,v,self)))$</td>
</tr>
<tr>
<td>$S_{i}::$</td>
<td>$T_{j}::$</td>
</tr>
<tr>
<td>$v = e &amp; SCond(self, v) \Rightarrow T_{j} \rightarrow \text{exists}(t \mid \text{TCond}(t) &amp; P(self,v,t))$</td>
<td>$TCond(self) \Rightarrow S_{i} \rightarrow \text{exists}(s \mid SCond(s,e) &amp; P^{-1}(s,e,self))$</td>
</tr>
<tr>
<td>$S_{i}::$</td>
<td>$T_{j}::$</td>
</tr>
<tr>
<td>$v = \text{Ref}[r.idTRef] &amp; SCond(self) \Rightarrow T_{j} \rightarrow \text{exists}(t \mid \text{TCond}(t) &amp; t_{rr} = v &amp; P(self,t))$</td>
<td>$w = \text{Ref}[r.idTRef] &amp; \text{TCond}(self) \Rightarrow S_{i} \rightarrow \text{exists}(s \mid SCond(s) &amp; s.r = w &amp; P^{-1}(s, self))$</td>
</tr>
</tbody>
</table>

$sq1 \rightarrow \text{intersection}(sq2)$ abbreviates $sq1 \rightarrow \text{select}(x \mid x : sq2)$ and hence retains the order of $sq1$. In some cases the inverse predicate is based on a functional inverse to the source-to-target mapping. For example, $t.g = \text{not}(s.f)$ can be trivially inverted to $s.f = \text{put}_{\text{not}}(s.f@\text{pre}, t.g)$ so that the view-update $s.f = \text{put}_{\text{not}}(s.f@\text{pre}, t.g)$ for $\text{get}_{\text{not}}$ defined as $\text{not}$ does not need to take account of the previous value $s.f@\text{pre}$ of $s.f$. The case of $t_{rr} = s.r \rightarrow \text{select}(P)$ is an example of a non-trivial view update $s.r = \text{put}_{\text{select}}(s.r@\text{pre}, t_{rr})$: given a value of $t_{rr}$, we determine an updated $s.r$ by taking the union of $s.r@\text{pre} \rightarrow \text{reject}(P)$ and $t_{rr} \rightarrow \text{select}(P)$. The first set are those elements of $s.r$ which are unaffected by changes to $t_{rr}$, whilst the second set consists of possibly new elements introduced by changes to $t_{rr}$. View updates are discussed in more detail in Section VI.

Table 5 summarises the inverse forms of different constraint kinds involving let variable definitions $v = E$, $v = \text{Ref}[id]$ or additional quantified variables $v : F$.²

²Let $v : T = e$ in $E$ is written as $v = e \iff E$ in our notation.

Implementation of the forward direction of $\tau$ is normally computed as $\text{stat}_{\text{LC}}(\text{Post}_{\tau})$: the UML activity derived from $\text{Post}_{\tau}$ when interpreted procedurally [35]. However, in order to achieve the correctness and hippocraticness bx properties, $\text{Inv}_{\tau}$ must also be considered: before $\text{stat}_{\text{LC}}(\text{Post}_{\tau})$ is applied to the source model $m$ and target model $n$, $m$ must be cleared of elements which fail to satisfy $\text{Inv}_{\tau}$. This is a case of the Cleanup before Construct pattern (Section IV).

The general cleanup constraint $\text{Cleanup}_{i}$ derived from $\text{Inv}_{\tau}$ is:

$$T_{j}::
\text{TCond}(\text{self}) \&
\text{not}(S_{i} \rightarrow \text{exists}(s | SCond(s) \& P_{j}(s, self))) \implies
\text{self} \rightarrow \text{isDeleted}()$$

This expresses that if a target instance $t$ does not satisfy $\text{Inv}_{\tau}(s,t)$ for any $s : S_{i}$, then $t$ is deleted.

However, this rule may delete more $T_{j}$ instances than are necessary: not only instances which correspond (i) to deleted $S_{i}$ instances, or (ii) to $S_{i}$ instances that no longer satisfy $SCond_{i}$, but also (iii) to $S_{i}$ instances satisfying $SCond_{i}$, but not $P_{j}(s, self)$. In this case, $\text{self}$ should not be deleted, instead $\text{stat}_{\text{LC}}(P_{j})$ should be performed to re-establish $P_{j}$ between the modified $s$ and $\text{self}$. Therefore, some form of source-target tracing is required, to distinguish these cases.
TABLE 6. Inverse of predicates (extended).

<table>
<thead>
<tr>
<th>$P(s, t)$</th>
<th>$P^\sim(s, t)$</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t = T_i[s.id]$</td>
<td>$s = S_i[t.id]$</td>
<td>idS primary key of $S_i$</td>
</tr>
<tr>
<td>$t, r \rightarrow TRef[s, r] \rightarrow \text{collect(idS)}$</td>
<td>$s, r \rightarrow SRef[t, r] \rightarrow \text{collect(idT)}$</td>
<td>idT primary key of $T_j$, $r, r$ ordered association ends</td>
</tr>
<tr>
<td>$t, r \rightarrow T\Ref[s, r] \rightarrow \text{idSRef}$</td>
<td>$s, r \rightarrow S\Ref[t, r, id\TRef]$</td>
<td>rr association end with element type $T\Ref$</td>
</tr>
<tr>
<td>$id\SRef$ primary key of $S\Ref$, $id\TRef$ primary key of $T\Ref$</td>
<td></td>
<td>$r$ association end with element type $S\Ref$</td>
</tr>
<tr>
<td>$T_j[s.idS], r = T\Ref[s, r, id\SRef]$</td>
<td>$S_i[t.idT], r = S\Ref[t, r, id\TRef]$</td>
<td>idS primary key of $S_i$, idT, id\TRef primary keys of $T_j$, $T\Ref$</td>
</tr>
<tr>
<td>$r$ has element type $S\Ref$, $rr$ has element type $T\Ref$</td>
<td></td>
<td>$r$ has element type $S\Ref$</td>
</tr>
<tr>
<td>$t, g \rightarrow T\Ref\Set(s) \rightarrow \text{closure}(r) \rightarrow \text{unionAll}(f, id\SRef)$</td>
<td>$s, f = S\Ref[t, g, id\TRef \rightarrow \text{oclIsUndefined}() \rightarrow \text{unionAll}(f, id\SRef)]$</td>
<td>$f$ many-valued, injective, $r$ : $S_i \rightarrow _S_i$ acyclic</td>
</tr>
<tr>
<td>$s, r \rightarrow \text{closure}(r) \rightarrow \text{unionAll}(f, id\SRef)$</td>
<td></td>
<td>$g : T_j \rightarrow _TRef, f : S_i \rightarrow _SRef$</td>
</tr>
</tbody>
</table>

We apply the Auxiliary Correspondence Model pattern to introduce String-valued identity attributes (primary keys) for source and target entities, to record and identify which source and target instances correspond: source instance $s$ corresponds to target instance $t$ when $s.sId = t.tId$. This also means that $S_i[t.tId] = s$ and $T_j[s.sId] = t$. By using the Entity Splitting and Entity Merging patterns, described in Section IV, one-many and many-one relations between source and target instances can also be recorded using this technique.

Tables 3 and 4 can be extended to include assignments to features based on object indexing lookups (Table 6).

If identity attributes have been introduced using Auxiliary Correspondence Model, the postconditions (rules) $C_i$ of $\tau$ have the typical form

$$S_i ::$$

$$SCond_i(self) \implies$$

$$T_i \rightarrow \text{exists}(t | t.tId = sId \& TCond_i(t) \& P_j(self, t))$$

and the corresponding invariant constraints $Inv_i$ are:

$$T_j ::$$

$$TCond_j(self) \implies$$

$$S_j \rightarrow \text{exists}(s | s.sId = tId \& SCond_j(s) \& P_j(s, self))$$

The cleanup constraints can then be simplified to:

$$T_j ::$$

$$TCond_j(self) \&$$

$$\text{not}(tId : S_i \rightarrow \text{collect}(sId)) \implies self \rightarrow \text{isDeleted}()$$

and

$$T_j ::$$

$$TCond_j(self) \& tId : S_i \rightarrow \text{collect}(sId) \&$$

$$\text{not}(SCond_j(S_i[tId])) \implies self \rightarrow \text{isDeleted}()$$

which correspond to cases (i) and (ii) above. This version is also potentially more efficient than the original cleanup constraint. The second case is omitted if $SCond_i$ is the default true. The updates of case (iii) are carried out in the $Post_i$ constraints. $\tau \rightarrow$ is defined to have postconditions consisting of the cleanup constraints, followed by $Post_i$, and hence it is an extension of $\tau$: its postconditions imply those of $\tau$. We denote by $\tau^\times$ the transformation that consists of the cleanup constraints. Then the forward transformation can be decomposed as the sequential composition of $\tau^\times$ and the original $\tau$: $\tau \rightarrow = \tau^\times; \tau$.

The cleanup constraints can be further optimized by writing them in the form:

$$T_j ::$$

$$TCond_j(self) \&$$

$$S_i[tId] \rightarrow \text{oclIsUndefined}() \implies self \rightarrow \text{isDeleted}()$$

and

$$T_j ::$$

$$TCond_j(self) \& \text{not}(S_i[tId] \rightarrow \text{oclIsUndefined}()) \&$$

$$\text{not}(SCond_j(S_i[tId])) \implies self \rightarrow \text{isDeleted}()$$

A single consolidated $Cleanup_i$ constraint is:

$$T_j ::$$

$$TCond_j(self) \implies$$

$$\text{if } S_i[tId] \rightarrow \text{oclIsUndefined}() \text{ then } self \rightarrow \text{isDeleted}() \text{ else if } \text{not}(SCond_j(S_i[tId])) \text{ then } self \rightarrow \text{isDeleted}() \text{ else true endif endif$$

A similar analysis and formation of invariants and cleanup constraints can be carried out for the other bx pattern cases of Table 2. Details are given in the appendix.
\( \tau^\leftarrow \) operates on both the source and target model, and propagates changes in the source model \( m \) to the target \( n \): if elements of \( m \) are deleted, so are the corresponding elements of \( n \). If elements of \( m \) are modified so that no rule applies to map an \( m \) element to a particular target element, then that target is deleted (this is the implicit deletion semantics of QVT-R [8], here made explicit by the \( \tau^\leftarrow \) rules). Otherwise, if a modified source element still corresponds by identity to a valid target element, that target is updated according to the forward rules. Finally, if a rule applies to map a source element (new or modified) to a target which does not already exist, an appropriate target is created. Changes to identity attribute values are not permitted.

**VI. INVERSE TRANSFORMATIONS AND VIEW UPDATES**

For the reverse transformation \( \tau^\rightarrow \), the roles of \( \text{Post}_t \) and \( \text{Inv}_t \) are interchanged: elements of the source model which fail to satisfy \( \text{Post}_t \) with respect to some element of the target model should be deleted or modified. The postcondition of the reverse transformation \( \tau^\rightarrow \) is \( \text{Post}_r \), the explicit form of \( \text{Inv}_t \). The key difference between \( \tau^\rightarrow \) and \( \tau^\leftarrow \) is that \( \text{Post}_r \) may involve view updates [4] of the source model, using implicit changes to source elements based on target data, because \( \tau \) is specified using explicit updates of the target based on the source.

If a predicate such as \( t.g = s.f \rightarrow \text{last}() \) or \( t.g = s.f \rightarrow \text{select}(P1) \) is inverted, the result is a predicate \( s.f \rightarrow \text{last}() = t.g \) or \( s.f \rightarrow \text{select}(P1) = t.g \). These are termed view updates, because they specify an update to \( s.f \) based on the required value of some view, function or selection of its data. There may not be a unique way to perform such an update.

We extend the definition of \( \text{stat}_L \)C so that the procedural interpretation \( \text{stat}_L \C(P) \) of a view update predicate \( P \) is a statement which makes \( P \) true by making minimal necessary changes to \( s.f \). using suitable \( \text{put}_E \) functions which take as input the previous value \( s.f @\text{pre} \) of \( s.f \) and the updated value \( t.g \) of the target feature: \( s.f := \text{put}_E(s.f @\text{pre}, t.g) \). This approach can be used to implement target-to-source change propagation for a bx where the target model is constructed from views of the source model.

Tables 7, 8, 9 show the extended definitions of \( \text{stat}_L \C \)C for some common view update predicates.

In the cases for \( \text{tail} \) and \( \text{front} \) in Table 7, \( d \) is the default element of the element type of \( f \). In the case for \( \text{collect} \), \( e^\leftarrow \) is an inverse to \( e \), defined for example, according to Table 3. Notice that in this case, elements are deleted from \( f \) if their \( e \) value is not in \( g \), then any extra elements \( y \) where \( y.e \) is in \( g \) but \( y \) is not in \( f @\text{pre} \), are added to \( f \). \( s \rightarrow \text{intersection}(t) \) is treated as for \( s \rightarrow \text{select}(x: t) \).

An assignment \( t.g = D(s.f \rightarrow \text{select}(P) \rightarrow \text{collect}(bldId)) \) can be inverted to \( s.f \rightarrow \text{select}(P) = B(t.g.bldId) \) if \( D \) is the corresponding target entity type for source entity \( B \) and similar for other cases of binary and unary collection operators in place of \( \text{select} \). The view update definitions can therefore also be used for bx that involve source-target correspondences using identity attributes.

**TABLE 7. View update interpretations for \( s \).**

<table>
<thead>
<tr>
<th>( P )</th>
<th>( \text{derived predicate} \text{stat}_L \C(P) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f \rightarrow \text{last}() = x )</td>
<td>( \text{if } f.size = 0 \text{ then } P := \text{Sequence}{x} \text{ else } f.size := x )</td>
</tr>
<tr>
<td>( f \rightarrow \text{front}() = g )</td>
<td>( \text{if } f.size &gt; 0 \text{ then } P := g \rightarrow \text{Sequence}{f.last} \text{ else } f := g \rightarrow \text{Sequence}{d} )</td>
</tr>
<tr>
<td>( f \rightarrow \text{first}() = x )</td>
<td>( \text{if } f.size = 0 \text{ then } P := \text{Sequence}{x} \text{ else } f[1] := x )</td>
</tr>
<tr>
<td>( f \rightarrow \text{tail}() = g )</td>
<td>( \text{if } f.size &gt; 0 \text{ then } P := \text{Sequence}{f.first} \rightarrow g \text{ else } P := \text{Sequence}{d} \rightarrow g )</td>
</tr>
<tr>
<td>( f \rightarrow \text{select}(P) = g )</td>
<td>( \text{if } f := f \rightarrow \text{select}(x: {x: g &amp; P } &amp; \text{not}(P)) \rightarrow \text{sequence-value f} ) ( (g \rightarrow \text{put}_E) )</td>
</tr>
<tr>
<td>( f \rightarrow \text{reject}(P) = g )</td>
<td>( \text{Same as } f \rightarrow \text{select}(\text{not}(P)) = g )</td>
</tr>
<tr>
<td>( f \rightarrow \text{collect}(e) = g )</td>
<td>( \text{if } f := \text{put}_E \rightarrow \text{reject}(x: x.e \notin g) \rightarrow {g \rightarrow \text{collect}(e^\leftarrow) \Rightarrow f \rightarrow \text{pre} } )</td>
</tr>
</tbody>
</table>

**TABLE 8. View update interpretations for \( e \).**

<table>
<thead>
<tr>
<th>( P )</th>
<th>( \text{derived predicate} \text{stat}_L \C(P) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x : a \rightarrow \text{intersection}(b) )</td>
<td>( x : a &amp; x : b )</td>
</tr>
<tr>
<td>( x : a \rightarrow b )</td>
<td>( x : a &amp; x : b )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{select}(P) )</td>
<td>( x : a &amp; x : P )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{reject}(P) )</td>
<td>( x : a &amp; x : \text{not}(P) )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{hasSet}(e) )</td>
<td>( x : a )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{hasSequence}(e) )</td>
<td>( x : a )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{sort}(e) )</td>
<td>( x : a )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{sortedBy}(e) )</td>
<td>( x : a )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{reverse}(e) )</td>
<td>( a := a \rightarrow \text{prepend}(x) )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{union}(b) )</td>
<td>( \text{if } a \rightarrow \text{union}(b) \rightarrow \text{includes}(x) \text{ then skip else } a := a \rightarrow \text{including}(x) )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{multiplicity a} )</td>
<td>( \text{if } a \rightarrow \text{union}(b) \rightarrow \text{includes}(x) \text{ then skip else if } a.size = 0 \text{ then } a := a \rightarrow \text{including}(x) )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{0.1-multiplicity a} )</td>
<td>( b := b \rightarrow \text{excluding}(x) )</td>
</tr>
</tbody>
</table>

**TABLE 9. View update interpretations for \( e^\leftarrow \).**

<table>
<thead>
<tr>
<th>( P )</th>
<th>( \text{derived predicate} \text{stat}_L \C(P) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x : a \rightarrow \text{intersection}(b) )</td>
<td>( x : a &amp; x : b )</td>
</tr>
<tr>
<td>( x : a \rightarrow b )</td>
<td>( x : a &amp; x : b )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{select}(P) )</td>
<td>( x : a &amp; x : P )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{reject}(P) )</td>
<td>( x : a &amp; x : \text{not}(P) )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{hasSet}(e) )</td>
<td>( x : a )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{hasSequence}(e) )</td>
<td>( x : a )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{sort}(e) )</td>
<td>( x : a )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{sortedBy}(e) )</td>
<td>( x : a )</td>
</tr>
<tr>
<td>( x : a \rightarrow \text{union}(b) )</td>
<td>( x : a \rightarrow \text{union}(b) )</td>
</tr>
</tbody>
</table>

An example of a case from Table 8 is the inverse mapping \( \text{person2family} \rightarrow \text{family2person} \). The \( \text{Post}^\rightarrow \) rules in this case have the form:

**Example:**

\( \text{Family}::=\text{exists}(fx) \left| fx.id = \text{familyId} \& \right.\)  
\( \text{FamilyMember}::=\exists(m) \left| m.memberId = \text{personId} \& \right.\)  
\( m : \text{fx.mother->union(fx.daughters)} \)
The update \( m : \text{fx.mother} \to \text{union}(\text{fx.daughters}) \) adds \( m \) preferentially to \( \text{fx.mother} \), only adding to \( \text{fx.daughters} \) if \( \text{fx.mother} \) is already set.

The cleanup constraints for \( \tau^\rightarrow \) are as follows in the general case:

\[
S_i \::= \
SCond_i(\text{self}) \& \text{not}(\text{slId} : T_j \to \text{collect}(\text{tlId})) \implies \text{self} \to \text{isDeleted}()
\]

for each \( C_i \), and

\[
S_i \::= \
SCond_i(\text{self}) \& \text{slId} : T_j \to \text{collect}(\text{tlId}) \& \text{not}(\text{TCond}_j(\text{tlId})) \implies \text{self} \to \text{isDeleted}()
\]

The latter is omitted if \( \text{TCond}_j \) is not present (ie., it is true).

As with the forward transformation, \( \tau^\leftarrow = \tau^\rightarrow \circ \tau^\rightarrow \), \( \text{stat}_{1}C(\text{Post}^\leftarrow) \) establishes \( \text{Inv} \), whilst the cleanup constraints together with \( \text{Post}^\leftarrow \) establish \( \text{Post} \). Thus \( \tau^\leftarrow \) satisfies correctness. Hippocraticness follows since neither the cleanup nor \( \text{Post}^\leftarrow \) constraints modify the source model if \( R \) already holds.

View updates can be defined for certain cases of user-defined functions by using syntactically-derived view update functions as in [34], [47], and [61]. In particular, any constraint predicate \( P(s, t) \) which computes a sequence-valued feature \( t.g = \text{get}(s.f) \) from sequence-valued \( s.f \) by using a recursively-defined function \( \text{get} \) on the elements of \( s.f \) can be inverted to \( s.f = \text{put}(s.f @ \text{pre}, t.g) \) for a recursively-defined view update function \( \text{put} \).

If \( \text{get} \) is schematically of the form:

\[
\text{get}(x) = \begin{cases} 
\text{Sequence}{} & \text{if } x.\text{size} = 0 \\
\text{put}(x.\text{first}, x.\text{tail}) & \text{if } P_1(x.\text{first}, x.\text{tail}) \\
\text{put}(x.\text{tail}, y.\text{tail}) & \text{else if } \ldots \\
\text{get}(x.\text{tail}) & \text{else get}(x.\text{tail}) \end{cases}
\]

Then a corresponding \( \text{put} \) is defined by:

\[
\text{put}(x, y) = \begin{cases} 
\text{Sequence}{} & \text{if } x.\text{size} = 0 \\
\text{put}(x.\text{first}, x.\text{tail}) & \text{if } P_1(x.\text{first}, x.\text{tail}) \\
\text{put}(x.\text{tail}, y.\text{tail}) & \text{else if } \ldots \\
\text{put}(x.\text{tail}, y.\text{tail}) & \text{else put}(x.\text{tail}, y.\text{tail}) \end{cases}
\]

where each \( \text{put}_i \) is an update function inverse to \( \text{get}_i \). Similarly for recursively-defined \( \text{get} \) functions from trees to sequences [62].

![Figure 8. UML2C transformation architecture.](image_url)

**VII. CASE STUDY: MAPPING UML TO C**

UML-RSDS already contains code generators for three versions of Java, and for C# and C++. If a code generator for a new language or language version is required, this can be written as a UML-RSDS transformation from the design metamodel (Figure 9 together with an OCL metamodel and activities metamodel) of UML-RSDS to the abstract and concrete syntax of the new target language, and then incorporated as a plugin into the UML-RSDS tools [44]. Here we consider the generation of ANSI C code from UML, and the extraction of a UML model from a C model. The code generator is factored into a model-to-model bx and a model-to-text transformation using the architecture of Figure 8.

The model-to-model bx \( \text{design}2\text{C} \) has source language the combined UML-RSDS metamodels for class diagrams, expressions and activities. Figure 9 shows the class diagram metamodel. The target language is a simplified version of the abstract syntax of C programs (Figure 10 shows the type and program structure parts of this metamodel).

We define \( \text{design}2\text{C} \) using the patterns of Table 2, to enable it to be used as the basis of a bx between UML and C. \( \text{design}2\text{C} \) consists of five subtransformations, mapping separately (i) types, (ii) class diagrams, (iii) expressions, (iv) activities, (v) use cases to C. Here we consider specifically the type mapping (i), and give extracts from (ii) and (iii).

**Table 10. Informal mapping between UML and C.**

<table>
<thead>
<tr>
<th>UML element e</th>
<th>C representation e’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class E</td>
<td>struct E ( ... );</td>
</tr>
<tr>
<td>Property p : T</td>
<td>member T’ p;</td>
</tr>
<tr>
<td></td>
<td>of the struct for p’s owner,</td>
</tr>
<tr>
<td></td>
<td>where T’ representis T</td>
</tr>
<tr>
<td>String type</td>
<td>char*</td>
</tr>
<tr>
<td></td>
<td>same named C types</td>
</tr>
<tr>
<td>int, long, double types</td>
<td>unsigned char</td>
</tr>
<tr>
<td>boolean type</td>
<td>struct E* type</td>
</tr>
<tr>
<td>Set(E) type</td>
<td>struct E**</td>
</tr>
<tr>
<td>Sequence(E) type</td>
<td>(array of E’ with no duplicates)</td>
</tr>
<tr>
<td>Operation op(p : P) : T of E</td>
<td>C operation</td>
</tr>
<tr>
<td></td>
<td>T’ op[E’ self, P’ p]</td>
</tr>
</tbody>
</table>

Table 10 shows the informal mapping between UML and C for classes, types and attributes.

Primitive types, entity types and collection types are successively mapped. \( \text{typeld : String} \) and \( \text{typeld : String} \) are new identity attributes introduced for Auxiliary Correspond-
dence Model. These provide a correspondence between the occurrences of types in the source UML-RSDS design and in the target C implementation. typeId is a string consisting of digits.

The postconditions mapping UML types to C are:

```plaintext
p.typeId = typeId & 
p.name = "int"
```

```plaintext
PrimitiveType::
name = "long" =>
CPrimitiveType->exists(p | p.typeId = typeId & p.name = "long")
```

```plaintext
PrimitiveType::
name = "double" =>
CPrimitiveType->exists(p | p.typeId = typeId & p.name = "double")
```

FIGURE 9. UML-RSDS class diagram metamodel.

FIGURE 10. C program and types metamodel.
p.ctypeId = typeId &
  p.name = "double"
)

PrimitiveType::
  name = "boolean" =>
  CPrimitiveType->exists( p |
    p.ctypeId = typeId &
    p.name = "unsigned char"
  )

PrimitiveType::
  name = "String" =>
  CPointerType->exists( t |
    t.ctypeId = typeId &
    CPrimitiveType->exists( p |
      p.name = "char" &
      t.pointsTo = p )
  )

Entity::
  CPointerType->exists( p |
    p.ctypeId = typeId &
    CStruct->exists( c |
      c.name = name &
      c.ctypeId = name &
      p.pointsTo = c )
  )

CollectionType::
  name = "Sequence" =>
  CArrayType->exists( a |
    a.ctypeId = typeId &
    a.duplicates = true &
    a.componentType = CType[elementType.typeId]
  )

CollectionType::
  name = "Set" =>
  CArrayType->exists( c |
    c.ctypeId = typeId &
    c.duplicates = false &
    c.componentType = CType[elementType.typeId]
  )

For this mapping we require that collection types cannot have collection element types. Classes correspond to structs by name, this is a distinct correspondence to that of UML and C types. The char type has ctypeId equal to the empty string.

In subtransformation (ii), the attributes (and association ends) owned by a class are mapped to members of its corresponding struct:

Entity::
  CStruct->exists( c |
    c.name = name &
    c.ctypeId = name &
    c.ctypeId = name &
    c.pointsTo = c )

The CStruct[name] instance is looked-up by the name primary key, and is not re-created, according to the Unique Instantiation for bx pattern. The CType instance corresponding to p.type for each property p is looked up by its ctypeId.

The type mapping constraints conform to the patterns of Table 2. In particular they include an example of the Entity Merging (horizontal) pattern: CPointerType instances can correspond either to String type instances or to entity type instances. These cases are distinguished in the C language model, enabling a valid inverse map to be defined. The mapping of entity types is an example of (vertical) Entity Splitting: a UML class (entity type) is mapped to a C pointer type and to a C struct. In this case these target instances both belong to the same class hierarchy, so they are given distinct ctypeId key values, both of which are derived 1-1 from the source instance.

Following the model-to-model transformation design2C, a model-to-text transformation genCtext produces C text from the C metamodel. For example:

CPrimitiveType::
  name = "int" =>
  PrimitiveType->exists( t |
    t.typeId = ctypeId &
    t.name = "int"
  )

CPrimitiveType::
  name = "long" =>
  PrimitiveType->exists( t |
    t.typeId = ctypeId &
    t.name = "long"
  )

CPrimitiveType::
  name = "double" =>
  PrimitiveType->exists( t |
    t.typeId = ctypeId &
    t.name = "double"
  )

CPrimitiveType::
  name = "unsigned char" =>
  PrimitiveType->exists( t |
    t.typeId = ctypeId &
    t.name = "unsigned char"
  )
t.typeId = cTypeId &
  t.name = "boolean" )

CPointerType::
p : CPrimitiveType & p.name = "char" &
  pointsTo = p =>
  PrimitiveType->exists( t |
    t.typeId = cTypeId &
    t.name = "String" )

CPointerType::
c : CStruct & pointsTo = c =>
  Entity->exists( e |
    e.typeId = cTypeId &
    e.name = c.name )

CArrayType::
duplicates = true =>
  CollectionType->exists( t |
    t.typeId = cTypeId &
    t.name = "Sequence" &
    t.elementType =
      CType[componentType.ctypeId] )

CArrayType::
duplicates = false =>
  CollectionType->exists( t |
    t.typeId = cTypeId &
    t.name = "Set" &
    t.elementType =
      CType[componentType.ctypeId] )

The inverse mapping from C members to UML properties
is (from Tables 3, 6):

CStruct::
  Entity->exists( e |
    e.name = name &
    members->forAll( m |
      Property->exists( p |
        p.name = m.name &
        p.isUnique = m.isKey &
        p.type = CType[m.type.ctypeId] &
        p : e.ownedAttribute ) ) )

Source-to-target change propagation is supported by
design2C^→. For example, if a type is changed from a
Set to a Sequence in a UML-RSDS model m, then the Post
constraint

CollectionType::
  name = "Sequence" =>
  CArrayType->exists( a |
    a.ctypeId = typeId &
    a.duplicates = true &
    a.componentType =
      CType[elementType.typeId] )

applies to set duplicates = true for the corresponding C type.

Other parts of the UML to C translation can also be
defined as a bx in the same manner: mappings from
(ii) class diagrams, (iii) expressions and (iv) activities to
corresponding C elements have been defined as part of a
code generator for C in the UML-RSDS tools version 1.8
(https://nms.kcl.ac.uk/kevin.lano/uml2web/). The mappings
(iii) and (iv) use a Recursive Descent style of transforma-
tion, due to the self-recursive structure of expressions and
activities. To invert such transformations, the Inverse Recur-
sive Descent pattern can be used (Figure 6), based upon the
Auxiliary Correspondence Model and Unique Instantiation
patterns for bx, with invariants and inverse rules defined in
a similar manner as for non-recursive rules. As an example,
for the mapping of expressions from OCL to C, we have the
following operation in the role of op in the pattern:

BinaryExpression::
  mapBinaryExpression(lexp : CExpression,
    rexp : CExpression) : CExpression
  pre:
    lexp = CExpression[left.expId] &
    rexp = CExpression[right.expId]
  post:
    CBinaryExpression->exists( c |
      c.operator =
        Expression.cop(operator) &
      c.left = lexp & c.right = rexp &
      c.type = CType[type.ctypeId] &
      result = c )

This has the inverse op^→ from C to OCL:

CBinaryExpression::
  mapCBinaryExpression(lexp : Expression,
    rexp : Expression) : Expression
  pre:
    lexp = Expression[left.cexpId] &
    rexp = Expression[right.cexpId]
  post:
    BinaryExpression->exists( e |
      e.operator =
        CExpression.uop(operator) &
      e.left = lexp & e.right = rexp &
      e.type = CType[type.ctypeId] &
      result = e )

where cop and uop are dual functions defining the cor-
responding operators in the two languages. For example,
the OCL operator and (&) corresponds to the C operator &&.

The complete transformation consists of over 250 rules and
operations, and is one of the largest UML-RSDS transforma-
tions that have been developed.

VIII. BX SPECIFICATION IN QVT-R

QVT-R [51] was designed as a declarative MT language
supporting (in principle) specification of bidirectional and
multi-directional transformations. However to achieve bidi-
rectionality, various restrictions need to be placed on QVT-R
specifications.

In particular, the use of recursive where invocations
between QVT-R relations can prevent the use of a transforma-
tion as a bx [46], as can the use of ‘black box’ imple-
mentations of relations [51]. In addition, the semantics in
the standard is incomplete and lacks formality. There is
a lack of guidance for the correct organization of specifi-
cations, in particular how data-dependencies and conflicts
between relations should be managed, and there is a lack of
guidance on the construction of specifications that support
bidirectional/multi-directional execution.

To address these problems, we have implemented a formal
and extended version of the logical semantics of [51] as a
mapping from QVT-R to UML-RSDS. The mapping makes explicit the semantic meaning of a QVT-R specification in a form that can be manually inspected or formally analyzed, including the use of formal proof.

A. QVT-RELATIONS (QVT-R)

QVT-R transformations are defined using relations, which are rules that relate elements of a source model to corresponding elements of a target model. For example the UML2C mapping from classes to C structs (and vice-versa) can be defined by:

```c
top relation Class2CStruct
{ enforce domain design e : Entity
  { name = n ;
  enforce domain C c : CStruct
  { name = n, ctypesId = n ;
}
}
```

Relations include at least one domain, each domain is associated to some model involved in the transformation. Domain root variables represent elements of the domain model, in this case an Entity instance e of the design model, and a c : CStruct instance of the C model. The keyword enforce indicates that the domain elements may be modified by the relation when executed in the direction of the model of the domain.

The relation maps an Entity instance to a CStruct instance when execution is performed in the C model direction. If instead the relation was executed in the design direction, a mapping from CStruct to Entity would result. The variable n : String is an auxiliary (local) variable of the relation, used to transfer feature values from one domain to the other.

Top relations correspond to use case postcondition constraints in UML-RSDS: they are applied to all source elements which satisfy their application conditions, in an arbitrary order. QVT-R also has non-top relations, which only execute if explicitly invoked from the where clause of a relation, and in this case the source elements are supplied as inputs to the called relation via the call parameters. Non-top relations correspond to update operations in UML-RSDS.

The UML2C mapping from classes to pointer types in QVT-R is:

```c
top relation Class2CPointerType
{ enforce domain design e : Entity
  { typeId = t ;
  enforce domain C p : CPointerType
  { ctypesId = t ;
}
}
```

As with UML-RSDS rules, QVT-R relations $R_j$ may depend upon the results of previously-executed relations $R_i$. In QVT-R these dependencies can be explicitly stated by a when $\{ R_i(pars) \}$ clause in $R_j$. For example, the following relation creates links between existing CPointerType and CStruct instances:

```c
top relation LinkCPointerTypeCStruct
{ enforce domain design e : Entity { }
  enforce domain C p : CPointerType
  { pointsTo = c : CStruct { } };
  when
  { Class2CStruct(e,c) and
    Class2CPointerType(e,p) }
}
```

This is an example of the Map objects before links pattern: the pointsTo link is only set once the objects e and p are available to link.

We refer to the domain root variables (e and p here) and the object template variables of their patterns (eg., c) as the object variables of the relation.

B. USING BX PATTERNS IN QVT-R

Most of the bx patterns described in Section IV can also be used in QVT-R, to define bx in the language:

- Structure Preservation: simple top relations such as Class2CStruct above, with no when clause and a single enforce domain for each model.
- Phased Construction: top relations $R$ which construct a target element $t$ and reference another top relation $R'$ in their when clause to look up a target element $t'$ which $R$ links to $t$.
- Entity Splitting (horizontal): two or more top relations which each map the same source domain $s : S_j$ to target elements of different concrete entity types; the relations have disjoint when conditions; (vertical): top relations with a single source domain $s : S_i$ and multiple (&gt; 1) target domains/variables.
- Entity Merging (horizontal): two or more top relations which map source domains of different entity types to a common target domain $t : T$, the top relations have disjoint when conditions; (vertical): top relations with multiple source domains/object variables and single target domains/variables.
- Map Objects before Links: a top relation does not create target elements, instead it links elements created by other relations, which are listed in its when clause. E.g., LinkCPointerTypeCStruct above.
- Lens: relation $R$ computes a target feature $t.s.f$ as a get function of source feature(s) $s.f$. The assignment $t.g = get(s.f)$ and its inverse $s.f = put(s.f@pre, t.g)$ must both be included in the where clause of $R$.
- Flattening/unflattening: hierarchical structure in the source model is discarded in the mapping to the target, possibly together with source elements. The reverse direction must reassemble the structure/elements.

The Auxiliary Correspondence Model pattern is implicitly provided by the trace mechanism of QVT-R. It can also be explicitly introduced by means of key declarations, specifying that designated attributes are identities. Cleanup before Construct is internally implemented by deletion actions (scheduled after creation/update actions [51]). Unique Instantiation is internally implemented by QVT-R’s check-before-enforce semantics.

---

3This use of @pre is supported by the QVT-R to UML-RSDS translation.
The same concepts of read and write frames \( rd() \), \( wr() \) from UML-RSDS can be adopted for QVT-R. A significant difference to UML-RSDS is that relations have a designated application direction, with some domains considered as sources and others considered as targets. Only features and entities of target domains can be written by the relation. In addition, elements bound in the \( \text{when} \) clause are assumed to already exist and are not created by the relation. Eg., for LinkCPointerType\( \text{CStruct} \) above, \( wr() \) in the \( \text{C} \) direction is just \((\text{CPointerType} :: \text{pointsTo})\), and in the \( \text{design} \) direction is empty (the relation has no effect).

Relations may use auxiliary variables to transfer data between domains, eg., the string-valued variable \( t \) in Class2CPointerType. These are considered internal to the relation, and – as with local variables in UML-RSDS – they are not included in its write and read frame.

**C. MAPPING QVT-R TO UML-RSDS**

For a QVT-R transformation \( \tau \) on distinct models \( s : SL \) and \( t : TL \) we can define corresponding unidirectional UML-RSDS transformations \( \tau^{\rightarrow} \) and \( \tau^{\leftarrow} \) for the two possible execution directions \( s \rightarrow t \) and \( t \rightarrow s \) of \( \tau \). Top relations are mapped to UML-RSDS rules, and non-top relations to UML-RSDS operations.

The enforce semantics of QVT-R ([51, Sec. B.2]) is based upon two predicates \( \text{create}(R, m) \) and \( \text{delete}(R, m) \) for each relation \( R \) and model \( m \) of the transformation. Enforcement of \( R \) in the \( m \) direction consists of applying \( \text{create}(R, m) \) to carry out necessary creation/update actions on \( m \) required by \( R \), and then \( \text{delete}(R, m) \) to delete elements of \( m \) which are not required to exist by \( R \).

In our mappings from QVT-R to UML-RSDS we give separate formal interpretations \( \text{Pres}_t(m), \text{Con}_t(m) \) and \( \text{Cleanup}_t(m) \) for the update, creation and deletion phases of a relational transformation \( \tau \) executed in the direction of model \( m \). As in the QVT-R to QVT-Core mapping of [51], trace entities \( \text{RStrace} \) are used to record when relations \( R \) on source domains \( s \) and target domains \( t \) have been successfully applied: \( \text{RStrace} \) has properties \( x : E \) for all the object variables \( x : E \) of \( R \). A tuple \((a, b)\) for source elements \( a \) and target elements \( b \) should be in the trace \( \text{RStrace} \) iff \( R \) has been successfully applied to \( a \) to update or create \( b \), and all these elements exist. These traces are tested when \( R(a, b) \) occurs as a rule call in a \( \text{when} \) clause.

For relation \( R \), \( \text{Pres}_t(m) \) is a UML-RSDS constraint that defines \( R \)'s change-propagation actions for elements that have already been matched (ie., by a previous execution of \( \tau \)), it is defined as

\[
\text{RStrace} ::
\begin{align*}
&\text{whenp} \land \bigwedge_{d \in sdom} \text{cpred}(d) \\
&\text{wherep} \land \bigwedge_{d \in tdom} \text{epred}(d)
\end{align*}
\]

\( \text{whenp} \) is the logical interpretation of the \( \text{when} \) clause, \( \text{wherep} \) of the \( \text{where} \) clause, \( \text{cpred}(d) \) of a non-target domain \( d \), \( \text{epred}(d) \) of a target domain \( d \) of \( m \), and the conjunction is taken over all non-target domains \( sdom \) in the antecedent (ie, domains with models \( m', m' \neq m \)) and over target domains \( tdom \) in the succedent (ie, domains with model \( m \)).

The scope of any \( \exists \) quantifiers in the \( \text{epred} \) formulae are extended over the remainder of the succedent. Details of \( \text{whenp}, \text{wherep}, \text{cpred} \) and \( \text{epred} \) are given in [36]. We collect all \( \text{Pres}_t(m) \) constraints for both top and non-top \( R \) into a UML-RSDS use case \( \text{Pres}_t(m) \) representing a first transformation phase \( \text{Pres}_t(m) \) of \( \tau \) executed in the \( m \) direction. This phase propagates element feature changes from source models to \( m \).

For example, \( \text{Pres}_{\text{Class2CStruct}}(C) \) is

\[
\text{Class2CStructStrace} :: \quad n = e.name \Rightarrow c.name = n \land c.ctypeld = n
\]

This constraint propagates name changes from \( \text{Entity} \) instances to \( \text{CStruct} \)s.

\( \text{Con}_t(m) \) deals with the case of creation of new \( m \) elements and \( R \) traces, for top relations \( R \). It only applies if the source elements \( s_1, \ldots \) are not already in \( \text{RStrace} \):

\[
\begin{align*}
&\text{whenp} \land \bigwedge_{d \in sdom} \text{cpred}(d) \\
&\text{not}(\text{RStrace} \rightarrow \exists (tr | \bigwedge_i tr.s_i = s_i)) \Rightarrow \bigwedge_{d \in tdom} \text{epred}(d) \land \text{wherep}
\end{align*}
\]

where \( \text{wherep} \) includes the creation \( \text{RStrace} \rightarrow \exists (tr | \bigwedge_i tr.s_i = s_i \land \bigwedge_j tr.t_j = t_j) \) of a new trace. Again, \( \exists \) quantifiers in the succedent apply over all of the succedent following their introduction. A predicate \( \theta_3(m) \) is formed as for \( \text{Con}_t(m) \) but without the predicates on \( \text{RStrace} \). The \( \text{Con}_t(m) \) constraints are placed in a use case representing a second \( \tau \) phase, \( \text{Con}_t(m) \). These constraints propagate element creation from source models to \( m \).

For example, \( \text{Con}_{\text{Class2CStruct}}(C) \) is:

\[
\begin{align*}
&e : \text{Entity} \land n = e.name \\
&\text{not}(\text{Class2CStructStrace} \rightarrow \exists (tr | tr.e = e)) \Rightarrow \text{CStruct} \rightarrow \exists (c.| c.name = n \land c.ctypeld = n \\
&\quad \quad \quad \quad \text{Class2CStructStrace} \rightarrow \exists (tr | tr.e = e \land tr.e = c ))
\end{align*}
\]

Unique instantiation/least change semantics is used for the operational interpretation of the \( \exists \) quantifier: if \( x : X \) already exists such that \( P \) holds for \( x \), then \( X \rightarrow \exists (x | P) \) in a succedent does not recreate \( x \). Target elements identified by a key value are looked-up by that value and re-used if they already exist. We use the \( \text{stat}_{\text{Ref}}(P) \) semantics of Section III-C for the least-change procedural interpretation of predicates \( P \).
The \( \text{Cleanup}_E(m) \) phase of \( \tau \) removes any spurious target elements which have not been produced by any relation. For target entity \( E \) of \( m \), \( \text{Cleanup}_E \) is defined as:

\[
E :: \not(r1\text{Strace} \rightarrow \exists \text{exists}(r1|1.1e = \text{self})) \& \ldots \& \not(rm\text{Strace} \rightarrow \exists \text{exists}(rm|rm\text{em} = \text{self})) \implies \text{self} \rightarrow \text{isDeleted}()
\]

where the \( rk, k = 1 \) to \( m \), are all the relations (top or non-top) in which the entity \( E \) occurs as the type of some target domain or target object template \( ek : E \). All \( \text{Cleanup} \) constraints are placed in a final \( \tau \) phase, \( \text{Cleanup}_E(m) \). Unlike in UML-RSDS, where the cleanup constraints precede the main transformation, in QVT-R the cleanup actions must follow the \( \text{Con}_E(m) \) phase because traces are explicitly constructed by that phase, instead of being implicit based on identity attribute values.

For example, \( \text{Cleanup}_{C\text{Struct}} \) is:

\[
\text{C\text{Struct}} :: \not(\text{Class2C\text{Struct}\text{Strace} \rightarrow \exists \text{exists}(tr| \text{tr.c} = \text{self}))} \& \not(\text{LinkC\text{PO\text{tr}\text{e}\text{Typ}\text{eC\text{Struct}\text{Strace} \rightarrow \exists \text{exists}(tr| \text{tr.c} = \text{self}))} \implies \text{self} \rightarrow \text{isDeleted}()
\]

The \( \text{Cleanup} \) constraints propagate element deletion from source models to \( m \), because if any element of a trace object is deleted, so is the trace.

Additionally, non-top rules \( R \) are interpreted as operations with postconditions defined as for \( \text{Con}_R(m) \). Details are in [36]. Regarding the \( \text{overrides} \) clause, we follow [64] in removing this by expanding out the definition of an overridden relation to include an additional \( \text{when} \) predicate which expresses that no overriding relation is applicable. A \( \text{when} \) test on an overridden abstract relation \( R_{(\text{pars})} \) is expanded to be a disjunction \( R_1(\text{pars}) \) or \( \ldots \) or \( R_n(\text{pars}) \) on all the concrete relations \( R_i \) overriding \( R \).

For each execution direction, the three phases are executed in the order \( \text{Pres}_E(m); \text{Con}_E(m); \text{Cleanup}_E(m), \) and together should establish the following as postconditions of \( \tau \): (i) that any existing target object \( tx : E \) of \( m \) must appear as a target property of some trace: \( E \rightarrow \text{forAll}(tx|1\text{Strace} \rightarrow \exists \text{exists}(r1|1.1r1 = tx)) \) or \( \ldots \) or \( \text{rmStrace} \rightarrow \exists \text{exists}(rm|rm\text{tm} = tx) \) where the \( ri \) are as for \( \text{Cleanup}_E \); (ii) that all trace objects in \( R\text{Strace} \) satisfy the logical relation \( \theta E(m) \) defined by \( R \) (ie., \( \text{Pres}_E(m) \) holds); (iii) for top relations \( R \), that any tuple of source elements that satisfy the antecedent \( \varphi_R \) of \( \theta E(m) \) appears in some \( R\text{Strace} \).

The following restrictions are needed on the QVT-R specification \( \tau \) in order that (i), (ii), (iii) are established by its logical interpretation: (a) No target features/entity names occur in the \( \text{when} \) clause or source domain patterns of any relation; (b) If two relations both write to the same target features or entities, these updates are non-conflicting (ie, they create/update disjoint sets of elements of the same target entity, or update disjoint sets of features of instances of the entity); (c) there are no self-conflicts of a relation \( R \), ie., one application of \( R \) cannot invalidate a different application; (d) for any mandatory target model association \( A^{\text{m}1 \rightarrow 1} B \), the \( a,r \) element for \( a : A \) must be created/linked to \( a \) by \( \tau \) if \( a \) is created by \( \tau \).
In this section we evaluate the efficiency of the bx produced by our two approaches, using the list reversal example of [63], identify differences in the expressiveness of the approaches, and summarise the advantages and disadvantages of the approaches.

The list reversal case is used by [63] to compare different implementation approaches for QVT-R. Figure 11 shows the metamodels of the case.

In the UML-RSDS version of this case study the name attributes of the list and element classes are assumed to be identity attributes. The forward rules are:

DoublyLinkedList ::
   DoublyLinkedList1 → exists(l).name = name

Element ::
   Element1 → exists(e).name = name & e.list = DoublyLinkedList1[1].name

Element ::
   e = Element1[name]  ⇒  e.target = Element1[source.name]

This uses Structure preservation, Phased construction and Map objects before links, and is automatically invertible according to Tables 5, 6.

In our QVT-R version, Map objects before links is used to remove dependence of element2element on itself:

top relation list2list
{ enforce domain forward
  fList : DoublyLinkedList
  { name = n };
  when { list2list(fList,rList) }
}
top relation element2element
{ enforce domain forward
  fElement : Element
  { target = fTarget : Element {} };
  enforce domain reverse
  rElement : Element1
  { source = rSource : Element1 {} };
  when
  { element2element(fElement,rElement)
    and
    element2element(fTarget, rSource) }
}

The QVT-R version does not need to assume that name attributes are identities for tracing purposes, but instead this restriction is needed in order to ensure absence of self-conflicts in the relations (if two different elements in the source list had the same names they would be mapped to a single target element by element2element, and link2link would then attempt to set conflicting links for this element).

Figure 12 shows the execution time in ms of the UML-RSDS version and the QVT-R (translated via UML-RSDS) version, using generated Java 4 code on a 32-bit JVM and single core of a 2.53GHz Intel i3 Windows 7 laptop. The UML-RSDS version has a low time complexity, and is comparable to the performance of optimised versions of the case presented in [63]. The QVT-R via UML-RSDS version is less efficient, because of its more complex trace management. We also show the execution time using Medini QVT (projects.ikv.de/qvt) to execute the QVT-R version. This uses a 32-bit JVM on a 2.6GHz Intel i7 machine. The execution time is substantially higher than the QVT-R via UML-RSDS implementation approach.

The expressiveness of UML-RSDS and QVT-R is similar, when restricted to follow bx pattern structures and the
guidelines (a) to (h) above for QVT-R. The QVT-R restriction (a) corresponds to a UML-RSDS constraint being type 1. Restriction (b) is semantic non-interference in a UML-RSDS transformation. Restriction (c) is internal consistency and confluence for a UML-RSDS rule.

A structure-preservation constraint $C_i$ using Auxiliary Correspondence Model

$$ S_i \triangleright \quad SCond_i(self) \quad \Rightarrow \quad T_j \rightarrow \exists t \mid t.1Id = sId \quad \land \quad TCond_j(t) \quad \land \quad P_j(self, t) $$

in UML-RSDS corresponds to a QVT-R relation:

```csharp
top relation mapSiToTj
{ enforce domain src s : Si
  { sId = id }{ SCond1(s) };
  enforce domain trg t : Tj
  { tId = id }{ TCond1(t) };
  where { Pj(s,t) }
}
```

executed in the $t$ direction. This is a structure preservation relation in QVT-R.

A difference in expressiveness in this case is that the $where$ clause in our extended QVT-R can also contain the dual predicate $P_j^-(s, t)$. This is not possible in UML-RSDS because there is no restriction of constraint write-frames based on execution direction.

A phased construction constraint has the same structure but $P_j(s, t)$ has the form $PCond_j(s, t) \land t.rr = TRef[s.r.idSRef]$ to assign a value to an association role end $t.rr$ based on prior mappings of $SRef$ and $TRef$ elements. In QVT-R this becomes:

```csharp
top relation mapSiToTj
{ enforce domain src s : Si
  { sId = id, r = sref : SRef {} } { SCond1(s) };
  enforce domain trg t : Tj
  { tId = id } { TCond1(t) };
  when { mapSRefToTRef(sref, tref) }
  where { PCondJ(s, t) }
}
```

This is a phased construction relation in QVT-R. Note that the mapping from $r$ to $rr$ is done element-by-element instead of by a single assignment in UML-RSDS.

An entity-splitting (vertical) constraint

$$ S_i ::
SCond_i(self) \quad \Rightarrow
T_k \rightarrow \exists t \mid t.1Id = sId \quad \land \quad TCond_k(t)
$$

where $TCond_k(t)$ has the form

$$ PCond_k(s, t) \quad \land \quad t.1Id = sId \quad \land \quad TCond_k(t) \quad \land \quad P_j(self, t, 1, 2))$$

corresponds to an entity-splitting (vertical) relation:

```csharp
top relation mapSiToTk1
{ enforce domain src s : Si
  { sId = id }{ SCond1(s) };
  enforce domain trg t1 : Tk
  { t1Id = id } { TCond1(t1) };
  enforce domain trg t2 : Tl
  { t2Id = id } { TCond2(t1, t2) };
  when { P(s, t1, t2) }
}
```

Likewise for horizontal splitting and for entity merging constraints.

A flattening constraint

$$ E ::
\begin{align*}
  & f : fs \land FCond(f) \quad \Rightarrow \\
  & G \rightarrow \exists g \\
  & \quad g.gId = f.fiId \land g.eId = eId \land \\
  & \quad GCond(g) \land P.self, f, g)
\end{align*}
$$

where $GCond(g)$ is a preceding rule that maps $g$ to $G$.

A map objects before links linking constraint

$$ S_i ::
T_j[idS_i].rr = TRef[r.idSRef]$$

corresponds to a linking relation in QVT-R:

```csharp
top relation linkTjTRef
{ enforce domain src s : Si
  { r = sref : SRef { idSRef = idref } } { SCond1(s) };
  enforce domain trg t : Tj
  { rr = tref : TRef { idTRef = idref } } { TCond1(t) };
  when { mapSiToTj(s, t) and 
  mapSRefToTRef{sref, tref}
}
```
inheritance. We can conclude that the expressiveness of the approaches is similar when restricted to the bx patterns and guidelines, and that usages of patterns in UML-RSDS translate into usages of the same patterns in QVT-R.

Table 11 compares the advantages and disadvantages of the two approaches we have presented in this paper. Whilst the UML-RSDS approach has advantages of efficiency and conciseness, the reverse transformation is derived as a secondary construct from the forward transformation. In QVT-R a single specification text describes both directions. In the Lens case this enables the specifier to write both put and get functions together, and to create new function pairs as necessary, whilst in UML-RSDS the put is looked-up in a fixed catalogue of put/get pairs.

**X. CONCLUSION**

We have defined a declarative approach for defining bidirectional transformations in UML, based on use cases with dual postcondition and invariant relations between source and target models. These use cases permit many-to-one and one-to-many bx, in addition to bijections. The definition of bx and their analysis is supported by the UML-RSDS tools [42].

By using standard UML notations and concepts, our approach should enable bx specifications to be retained and maintained independently of particular transformation technologies. The use of OCL for transformation specification facilitates the automated derivation of transformation invariants and reverse rules. Translations from ATL and ETL to UML-RSDS have been developed [37], [40], and these potentially enable our bx approach to also be applied to unidirectional transformations in these languages. Our approach scales up to transformations of practical size, and we have given extracts from a UML to C code generator with over 250 mapping rules and operations.

We have also described an approach using QVT-R, utilising a translation from QVT-R to UML-RSDS to extend the expressiveness of QVT-R, and to provide efficient execution of different directions of the QVT-R specifications. We have described how bx patterns can be used in QVT-R to support the definition of bx in the language.

**APPENDIX**

**CORRECTNESS AND HIPPOCRATICNESS PROPERTIES FOR \( \tau \rightarrow \)**

In this section we give a formal justification that bx defined using the UML specification approach described in Section V do satisfy the bx properties of correctness and Hippocraticness, and that they satisfy local Hippocraticness and least change principles subject to some restrictions.

We consider the case of UML-RSDS transformations \( \tau \) that satisfy the patterns of Table 2, with source language entities \( S_i \) and target language entities \( T_j \), and whose constraints are of type 1, satisfying syntactic non-interference. Identity attributes are used to implement Auxiliary Correspondence Model. This means that \( \tau \) is a separate-models transformation with source language \( SL \) and target language \( TL \), and postcondition \( Post_\tau \) is an ordered conjunction of constraints \( C_j \) each of the form:

\[
S_i :: \quad SCond_i(self) \implies T_j \rightarrow exists(t | t.tId = sId \& TCond_j(t) \& P_j(self, t))
Inv_\tau \text{ is therefore a conjunction of constraints } Inv_\tau \text{ of the form }
\]

\[
T_j :: \quad TCond_j(self) \implies S_i \rightarrow exists(s | sId = s.sId \& SCond_i(s) \& P_j(s, self))
\]

A consequence is that \( S_i \rightarrow select(SCond_i) \) and \( T_j \rightarrow select(TCond_j) \) are isomorphic, with the isomorphism given by mapping \( s : S_i \rightarrow T_j[s.sId] \).

The simplified cleanup constraints for \( \tau \rightarrow \) are therefore:

\[
T_j :: \quad TCond_j(self) \& S_i[tId] \rightarrow oclIsUndefined() \implies self \rightarrow isDeleted()
\]

and

\[
T_j :: \quad TCond_j(self) \& tId : S_i \rightarrow collect(sId) \& not(SCond_i(S_i[tId])) \implies self \rightarrow isDeleted()
\]

The second constraint is omitted if \( SCond_j \) is absent (ie., \( true \)). We denote the collection of the cleanup constraints by \( Cleanup \). These are the postconditions of \( \tau \rightarrow \). Bx correctness holds, since the cleanup constraints together with \( Post_\tau \) establish \( Inv_\tau \); the cleanup constraints remove any target model elements that fail to correspond by identity to source model elements in the domain of \( \tau \), and \( Post_\tau \) modifies target model elements that do correspond to valid source elements, in order to re-establish \( Inv_\tau \). By construction, \( statsC(Post_\tau) \) preserves \( Inv_\tau \) and establishes \( Post_\tau \). Therefore, the sequential composition of
As for entity merging, there are two versions of this pattern: (i) two or more separate postconditions on $S_i$, with disjoint $SCond_i$ conditions, mapping to distinct target entity types $T_k$, $T_j$; (ii) a single constraint which maps one $S_i$ instance to multiple linked instances of different $T_j$. The first case is the reverse of Entity Merging (horizontal), and can be treated in a similar way to the Structure Preservation case, provided that syntactic or semantic non-interference of $Post_{i}$ holds. The second case is common in refinement transformations, and in some migrations. Constraints, eg., $C_1$, will have the form

$$S_1 : \begin{align*}
SCond_1(self) & \implies T_1 \rightarrow \exists t | t.t1ld = s1ld \land
\quad TCond_1(t) \land P_1(s, t) \end{align*}$$

and

$$S_j : \begin{align*}
SCond_j(self) & \implies T_j \rightarrow \exists t | t.t1ld = s2ld \land
\quad TCond_j(t) \land P_j(s, t) \end{align*}$$

Provided that the respective $TCond_i$, $TCond_j$ conditions are pairwise disjoint, and the sets of identities of $S_i \rightarrow select(SCond_i)$ and $S_j \rightarrow select(SCond_j)$ are disjoint, the same construction and argument for the bx correctness apply as in the Structure Preservation case. The invariants are

$$T_k : \begin{align*}
TCond_k(self) & \implies S_i \rightarrow \exists t | s.s1ld = t1ld \land
\quad SCond_i(s) \land P_i(s, self) \end{align*}$$

and

$$T_k : \begin{align*}
TCond_k(self) & \implies S_j \rightarrow \exists t | s.s2ld = t1ld \land
\quad SCond_j(s) \land P_j(s, self) \end{align*}$$

and hence $S_i \rightarrow select(SCond_i)$ is isomorphic to $T_k \rightarrow select(TCond_k)$ and $S_j \rightarrow select(SCond_j)$ is isomorphic to $T_k \rightarrow select(TCond_k)$. Semantic non-interference of $Post_j$ holds, because the set of $T_k$ instances produced from $S_j$ is disjoint from the set produced from $S_j$, although syntactic non-interference fails. Cleanup constraints can be derived from the invariants as for the previous cases. The vertical variant is treated similarly.

Entity Splitting (Section IV) involves one source entity type mapping to two or more different target entity types. As for entity merging, there are two versions of this pattern: (i) two or more separate postconditions on $S_i$, with disjoint $SCond_i$ conditions, mapping to distinct target entity types $T_k$, $T_j$; (ii) a single constraint which maps one $S_i$ instance to multiple linked instances of different $T_j$. The first case is the reverse of Entity Merging (horizontal), and can be treated in a similar way to the Structure Preservation case, provided that syntactic or semantic non-interference of $Post_j$ holds. The second case is common in refinement transformations, and in some migrations. Constraints, eg., $C_1$, will have the form

$$S_1 : \begin{align*}
SCond_1(self) & \implies T_1 \rightarrow \exists t | 1.r1ld = s1ld \land
\quad TCond_1(t) \land P_1(s, t) \end{align*}$$

$$T_2 \rightarrow \exists t | 2.r2ld = s2ld \land TCond_2(t) \land P_2(s, t)$$

Both $t1$ and $t2$ are given $id$ values $self.sld$ (in the case that $T_1$ and $T_2$ are not disjoint, distinct identifier values derived 1-1 from source instances should be used instead). This should be the only constraint in the transformation that creates $T_2$ instances. Other constraints can create $T_1$ instances if their $TCond_j$ conditions are disjoint from $TCond_i$. $Inv_j$ is of the form

$$T_1 : \begin{align*}
TCond_1(self) & \land t1 : T_2 \land t2.r2ld = t1ld \land
\quad TCond_2(t2, self) \implies S_1 \rightarrow \exists t | 1ld = s1ld \land SCond_1(s) \land
\quad P_1(s, self, t2) \end{align*}$$

The $t2 : T_2$ expression acts like an additional $T_2 \rightarrow forall(t2 \ldots)$ quantifier over the remainder of the constraint. The conjunction $t2 : T_2 \land t2.r2ld = t1ld$ can be optimised to $t2 = T_2[t1ld]$. For this type of rule, $S_1 \rightarrow select(SCond_1)$ is isomorphic to $T_1 \rightarrow select(TCond_1)$, however parts of the data of elements $s : S_1$ may be represented in the $t2 : T_2$ instances linked to the $t1 : T_1$ instance corresponding to $s$, rather than directly in $t1$.

The cleanup constraints for such constraints are:

$$T_1 : \begin{align*}
TCond_1(self) & \land t2 = T_2[t1ld] \land
\quad TCond_2(t2, self) \land
\quad not(S_1 \rightarrow \exists t | 1ld = s1ld \land SCond_1(s)) \implies
\quad self \rightarrow isDeleted() \land t2 \rightarrow isDeleted() \end{align*}$$
The effect of such a bx is that a ‘cluster’ of target instances (such as t₁, t₂) are related to each source instance, with all elements of the cluster having identities based on the source instance identity value. An example in UML2C is the mapping of a UML class to a CStruct and a CPointerType.

Again, statLC(Postₜ) establishes Postₜ and preserves Invᵣ, whilst statLC(Cleanup) together with statLC(Postₜ) establishes Invᵣ. Thus correctness follows. Hippocraticness holds because target model pairs (or clusters) t₁, t₂, . . . are only deleted if they do not correspond to a source instance. Unique Instantiation can apply to multiple exists quantifiers in the same manner as to single quantifiers. Local hippocraticness applies to a source entity type instance and its corresponding target cluster. The principle of least change follows from the target model – for elements that the operation is applied to. The specifier should ensure that the operation is always invoked on all instances of E satisfying ECond.

For the Map Objects before Links pattern (Figure 2), the construction of target instances is separated from the linking of these instances. In the case that Sᵢ corresponds to Tᵢ and SRef to TRef via one of the other patterns considered above, the linking constraints have the form:

\[ Sᵢ :: Tᵢ[idSᵢ],rr = TRef[r.idSRef] \]

These define target model association ends rr from source model association ends r, looking-up target model elements Tᵢ[idSᵢ] and TRef[r.idSRef] which have already been created by preceding constraint(s). The linking constraints can be inverted to a dual form which defines source data from target data as:

\[ Tᵢ :: Sᵢ[idTᵢ],rr = SRef[rr.idTRef] \]

according to Table 6.

The invariant for the linking constraint is:

\[ Sᵢ :: Tᵢ[idSᵢ],rr ≅ TRef[r.idSRef] \]

That is, every element of t.rr should correspond to an element of s.r, if s and t correspond.

Cleanup constraints are not needed for linking constraints, because deletion of SRef instances is propagated to deletion of TRef instances by the cleanup constraint for TRef construction (destruction of an instance includes removing it from every association end in which it resides, in UML-RSDS semantics). Other modifications to r are propagated to rr by the linking constraint itself.

By a similar argument to the previous cases, bx correctness, hippocraticness and least change can be shown. The re-establishment of Postₜ and Invᵣ for corresponding source and target elements is now performed by combination of the construction and linking Postₜ constraints.

REFERENCES


Action Language for Foundational UML (ALF), V1.0.1, OMG, 2015.

L. Paolini, M. Piccolo, and L. Roversi, “A class of reversible primi-

F. Hermann et al., “Model synchronization based on triple graph gram-


F. Hermann et al., “Model synchronization based on triple graph gram-

T. Hettel, M. Lawley, and K. Raymond, “Model synchronization: Defini-


Z. Hemel, L. C. L. Kats, D. M. Groenewegen, and E. Visser, “Code genera-


D. S. Kolovos, R. F. Paige, and F. A. C. Polack, “The Epsilon transforma-


K. Lano and S. Kolahdouz-Rahimi, “Model-transformation design pat-


KEVIN LANO has worked for over 25 years in the fields of system specification and verification. He was one of the originators of model-driven engineering, and has been a leading advocate of improving the precision of software modeling, and in applying software engineering principles to transformation construction. In recent years, he has worked on the integration of model-based development and agile development.

SHEKOUFEH KOLAHDOUZ-RAHIMI received the Ph.D. degree in computer science from Kings College London, in 2013. She is an Assistant Professor with the Computer Engineering Department, University of Isfahan, where she is an Active Member of the Model Driven Software Engineering Research Group. Her research interests include design patterns for model transformation, specification and verification of model transformations, bidirectional model transformations, domain-specific modeling languages, and search-based software engineering.

SOBHAN YASSIPOUR-TEHRANI has been working on requirements engineering (RE) in model transformations (MT). He has developed a RE technique suitability framework for MT projects by which the developer(s) are able to select the most appropriate RE technique for a particular requirement or a set of requirements. His research interests include RE, MT, goal modeling, and agile software development.

* * *