Co-Evolving Code with Evolving Metamodels

Djamel Eddine Khelladi  
CNRS, IRISA, Univ. Rennes  
Rennes, France  
djamel-eddine.khelladi@irisa.fr

Benoit Combemale  
Université Toulouse & Inria Rennes  
Rennes, France  
benoit.combemale@irisa.fr

Mathieu Acher  
Univ. Rennes, Inria, IRISA  
Rennes, France  
mathieu.acher@irisa.fr

Olivier Barais  
Univ. Rennes, Inria, IRISA  
Rennes, France  
olivier.barais@irisa.fr

Jean-Marc Jézéquel  
Univ. Rennes, Inria, IRISA  
Rennes, France  
jean-marc.jezequel@irisa.fr

ABSTRACT

Metamodels play a significant role to describe and analyze the relations between domain concepts. They are also cornerstone to build a software language (SL) for a domain and its associated tooling. Metamodel definition generally drives code generation of a core API. The latter is further enriched by developers with additional code implementing advanced functionalities, e.g., checkers, recommenders, etc. When a SL is evolved to the next version, the metamodels are evolved as well before to re-generate the core API code. As a result, the developers added code both in the core API and the SL toolings may be impacted and thus may need to be co-evolved accordingly. Many approaches support the co-evolution of various artifacts when metamodels evolve. However, not the co-evolution of code. This paper fills this gap. We propose a semi-automatic co-evolution approach based on change propagation. The premise is that knowledge of the metamodel evolution changes can be propagated by means of resolutions to drive the code co-evolution. Our approach leverages on the abstraction level of metamodels where a given metamodel element has often different usages in the code. It supports alternative co-evaluations to meet different developers needs. Our work is evaluated on three Eclipse SL implementations, namely OCL, Modisco, and Papyrus over several evolved versions of metamodels and code. In response to five different evolved metamodels, we co-evolved 976 impacts over 18 projects. A comparison of our co-evolved code with the versioned ones shows the usefulness of our approach. Our approach was able to reach a weighted average of 87.4% and 88.9% respectively of precision and recall while supporting useful alternative co-evolution that developers have manually performed.

1 INTRODUCTION

Model-Driven Engineering (MDE) has proven to be effective in the development and maintenance of large scale systems [20, 21]. Notably, by easing the task of building Software languages (SLs) and their tooling that can play a significant role in all phases of development processes [46].

When comes the time to specify a SL, the metamodel is pivotal and a central artifact [21]. It introduces the domain concepts, which are later used as a cornerstone to build the expected tooling (e.g., editor, checker, compiler, data access layers, etc.). Regular Expressions, CSS, GraphViz are examples of SLs that use an explicit initial specification to later drive the development of the corresponding tooling [10]. However, it is often overlooked that metamodels are inputs for complex code generators that leverage on the abstractions used in metamodels (e.g., bidirectional references, containment, ...) to generate complex and industrial code that may include advanced architecture and different ways to manipulate instances.

As a prominent example, Eclipse Modeling Framework (EMF) [45] supports the generation of Java code consisting of a core API (both generic and domain-specific), adapters, serialization facilities, and an editor, all from metamodels. This code can be used to load and manipulate the model instances. More importantly, it is often enriched by developers to offer additional functionalities and on top of which the SLs’ toolings are further developed.

Today, many Eclipse projects\(^1\) leverage on the code generated by EMF to support developers in various tasks (e.g., developing advanced editors, specific model transformations, etc.). As a result, many languages have been developed using EMF and its ecosystem, such as BPMN [41] or UML [43]. In all these examples, the initial core API generated with EMF have been complemented, either manually or through other tools from the EMF ecosystem, to provide advanced functionalities and tooling such as simulation, debugging, etc. A metamodel and its corresponding implementation are hence a cornerstone when building a SL and its tooling.

One of the foremost challenges to deal with in MDE is the evolution of metamodels and its impact on artifacts that use metamodels as a foundation. As a result of a metamodel evolution, the core API is re-generated and may impact the

---

\(^1\) Cf. https://marketplace.eclipse.org

ICSE ’20, May 23–29, 2020, Seoul, Republic of Korea

ACM Reference Format:

additional code implemented by developers. Hence, it may need to be co-evolved accordingly. Knowing that other projects/SLs can depend on the impacted code, it is essential to efficiently co-evolve it whenever the metamodel evolves. In the last decade, the co-evolution challenge has been extensively addressed in MDE. However, the literature so far has focused on the co-evolution between metamodel and models [5, 11, 18, 24, 25, 49], constraints [2, 6, 27, 33], and transformations [12, 13, 23, 32, 34]. To the best of our knowledge, there exist no work that aims at supporting the co-evolution between metamodels and code.

This paper contributes in filling this gap. We propose a semi-automatic approach to co-evolution of code when the metamodels evolve. The co-evolution propagates the metamodel changes to the code. Change propagation showed to be efficient in many domains, such as in the context of artifact recommendation or co-evolution of models or transformations [7, 9, 32]. This paper leverages on the knowledge provided by the metamodel changes during evolution which are propagated by means of code co-evolution. For example, knowing that a property \( p \) in the metamodel is moved from one class to another through a reference \( \text{ref} \), it is likely that the navigation path to \( p \) will also be updated in the code by either extending it with \( \text{ref} \) or reducing it. The current approach first runs an impact analysis to identify all impacts at the code level.

Our approach leverages on the abstraction level of metamodels where a given metamodel element has often a complex implementation in the code with various usages of such a code. For example, for a single metamodel class, an interface and a class are generated in the code, but also a method creation in the factory class, etc. Our impact analysis is designed to identify all impacted usages of a given changed metamodel element. Then, we propose a catalog of resolutions that is used as a basis for our code co-evolution in response to metamodel evolution. Developers can select among our supported alternative resolutions.

Our approach with the proposed resolutions are evaluated on three Eclipse EMF-based SL implementations, namely OCL [37], Modisco [36], and Papyrus [38], which have been developed for more than 10 years and have been evolved several times. We collected projects consisting of both original and evolved versions of metamodels and code. Thus, we evaluated to what extent can our approach correctly co-evolve the projects’ code in response to the metamodel evolutions. Within our case studies, 477 changes were considered on five metamodels, which caused 976 impacts on 18 projects. When comparing our performed co-evolution to the expected co-evolution from the original to the evolved versions of the projects, our approach showed to support useful alternative co-evolutions that developers have manually performed. Our approach reached a weighted average (by the number of metamodels changes) of 87.4% and 88.9% respectively of precision and recall of co-evolution. Our approach also showed to be useful by maintaining impacted code while it was unnecessary deleted.

This paper is structured as follows. Section 2 introduces a background and a motivating example. Section 3 presents our overall approach. Section 4 details our evaluation results Section 5 threats to validity. Section 6 discusses related work and how our work distinguishes from code migration approaches. Finally, Section 7 concludes the paper and reflects on future works.

2 BACKGROUND AND MOTIVATING EXAMPLE

This section first gives a background on the role of metamodels when building SLs and their toolings and discusses the scenarios of co-evolution that arises between metamodels and code. After that, it introduces an example taken from the OCL case study in Eclipse of metamodel and code co-evolution to motivate our work.

2.1 Background

Figure 1 depicts an SL structure and usage as in practice, for instance, as in the Eclipse platform. Metamodels are cornerstones when building a SL and its toolings [14, 47]. Metamodels define the aspects of a business domain, i.e. the main concepts, their properties, and relationships between them [3].

Once the metamodels are carefully specified, working piece of code is generated consisting of a core API [45], which is often enriched with additional code to offer more advanced functionalities. For instance, methods defined in classes in the metamodel, are generated without their bodies, which developers must implement. This part consists of enriching the generated core API. Developers also integrate additional classes to implement advanced functionalities (language services, language tooling, ...). This code is built on top of the generated core API.

Take for example the Regular Expression SL. From its metamodel the generated core API allows to parse a given regular expression and to manipulate it problematically (navigate in the AST, modifies it, etc.). However, no advanced functionalities are provided, such as the interpretation and evaluation of the given regular expression. This is the tooling that developers will add on top of the core API. Both depend and uses the generated core API from the metamodel, as shown in Figure 1.

When the metamodel evolves, the same part of the generated code can be re-generated again. As a consequence, the code manually integrated by developers may be impacted and may need to be co-evolved accordingly as well. The next section illustrates these challenges on a real-world use case.

2.2 Motivating Example

To motivate our work, this section introduces an example of metamodel and code co-evolution taken from the implementation of the OMG standard Object Constraint Language (OCL) [37, 42] in the Eclipse platform, which has evolved several times in the last years. OCL is a declarative language based on first-order logic, to define constraints on models, i.e.,
conditions that must hold on the models. Part of the OCL implementation is the OCL Pivot which provides the unified merged OCL functionality for the Ecore or UML metamodel instances. The OCL Pivot is implemented as an SL with a metamodel consisting of 111 classes in the version of 3.2.2.

Figure 2 depicts a small excerpt of the Pivot metamodel (6 out of 111 metaclasses). It illustrates the domain concepts Property, MultiplicityElement, Model, NamedElement, Namespace and ClassifierType. From all of these metaclasses, a first code is generated (i.e., core API), consisting of interfaces, classes, a factory, a package, etc. Listing 1 shows a snippet of the generated interfaces and classes for the metamodel excerpt in Figure 2 for the OCL Pivot SL. It is further enriched by the developers with a validation functionality in the new class PivotValidator, as shown in Listing 2. Another SL, namely OCL Base, also leverages on the OCL Pivot and its generated interfaces and classes in a conversion functionality from OCL Pivot to OCL Base, as shown in Listing 3.

In version 3.4.4, the OCL Pivot metamodel evolved significantly with various changes, among which, developers: 1) deleted the metaclasses MultiplicityElement and Model along with their properties, 2) renamed the method validateCompatibleInitialiser in the class Property to validateCompatibleDefaultExpression, 3) renamed the class ClassifierType to MetaClass, and 4) pushed the property ownedRule from NamedElement to Namespace. Naturally, the Code of Listing 1 was re-generated from the evolved version of the metamodel, which impacted the existing code depicted in Listings 2 and 3. The different caused impacts are underlined in red and their co-evolution is underlined in green in Listings 4 and 5 corresponding to version 3.4.4.

For some impacts, the co-evolution can be unique. For example, the renamed method validateCompatibleInitialiser to validateCompatibleDefaultExpression or the renamed class ClassifierType to MetaClass are propagated by developers to the code level as well and renamed in the Java class PivotValidator (lines 2 and 6 in Listing 4). However, some other impacts can be co-evolved in many different ways. Indeed, in the class Pivot2CSConversion, due to the deleted properties lower, upper, isOrdered, isUnique in the class MultiplicityElement, two different co-evolutions were applied. The navigation paths to the methods getLower() and getUpper() were deleted without deleting the whole variable declaration statement (lines 7 and 8 in Listing 3). Whereas, the whole statement of the method call refreshQualifiers is deleted rather than deleting the navigation paths to the methods getIsOrdered() and getIsUnique() (lines 12 and 13 in Listing 3). In the class PivotValidator the import and the whole method is deleted due to the delete of MultiplicityElement (lines 3 and 12 in Listing 2). Here we already see that a delete change in the metamodel can be co-evolved in different manners to cover different possible co-evolutions to different parts of the code.

Similar situation can even occur for the same impacting change. Take the pushed property ownedRule from NamedElement to Namespace. In the method refreshClassifier, the call to getOwnedRule() is replaced by getOwnedInvariant() (line 17 in Listing 3). Whereas, for the same impacted call to getOwnedRule() in the method refreshTypedElement, its whole statement of the method call visitDeclarations is deleted (line 22 in Listing 3). Other possible co-evolution could also be to introduce a cast or a cast with an if test of the type. This emphasizes the fact that developers should be involved in the co-evolution process to decide which alternatives should be applied.

Listing 1: Excerpt of the generated code in the SL org.eclipse.ocl.examples.pivot.

```java
1 //MultiplicityElement Interface
2 public interface MultiplicityElement extends Element { ... }
3 //Property Interface
4 public interface Property extends Feature, ... { ...
5 ... } 6 boolean validateCompatibleInitialiser(
6     DiagnosticChain diagnostics, Map<Object, Object> context);
7 ... 8 }
9 //PropertyImpl Class
10 public class PropertyImpl implements Property {
11 ... 
```
3 APPROACH

This section presents our approach of code co-evolution with evolving metamodels. Figure 3 shows our overall co-evolution process of code. It first identifies the metamodel changes and then runs an impact analysis on the code [1]. Here, we bridge the gap between the two abstraction levels of metamodels and code. Then a change propagation-based co-evolution is applied [2]. The developer is involved throughout the co-evolution to decide among alternatives, similarly as developers would choose among quick fixes in today’s IDEs to fix code errors. Finally, the chosen resolutions are applied to co-evolve the code [3].
3.1 Change Detection and 1 : n Impact Analysis

Before we start co-evolving code, we run an impact analysis to detect the code parts that must be co-evolved.

The starting point of our impact analysis is the changes applied during the metamodel evolution. Numerous approaches exist in the literature, such as [4, 29, 30, 35, 48, 50]. To not restrict our approach to a particular detection approach, we define a Change interface (see Figure 3) that serves to bridging existing change detection approaches with our internal knowledge of how the core API code is generated to cover those impacts on the code to co-evolve. We leverage on the power of abstraction of the metamodels [28] to identify different usages. Our change interface considers both atomic and complex changes [19]. Atomic changes are additions, removals, and updates of a metamodel element. Complex changes consist in a sequence of atomic changes combined together. For example, push property is a complex change where a property is moved from a superclass to its subclasses. This is composed of two atomic changes: delete a property and add a property.

With the detected metamodel changes, we run an impact analysis [22] on the code. To this end, we parse the code to access the Abstract Syntax Tree (AST). We then establish a mapping between the metamodel elements $e_1, \ldots, e_n$ and the code AST nodes corresponding to $e_1, \ldots, e_n$, as illustrated in Algorithm 1.

In contrast to the co-evolution of metamodel and models(transformations/constraints) that considers a 1 : 1 relationship between the metamodel elements and their use in the impacted artifacts. Our impact analysis considers a 1 : n relationship where a metamodel element is used in n different forms at the code level.

Indeed, take for example the class Property in the OCL Pivot metamodel. At the code level, an interface Property and a class PropertyImpl are generated. But also in the classes PivotFactoryImpl and PivotPackageImpl, respectively the methods createProperty() and getProperty() are generated. All this n code correspondences of the class Property are identified in Algorithm 1 (Lines 5-11). Other metamodel elements also have n code correspondences, such as attributes. For example, for the attribute implicit in the class Property, a getter and a setter getImplicit()/setImplicit() are generated, but also the method getImplicit() is generated in the class PivotPackageImpl (handled in Algorithm 1 (Lines 13-20)).

Therefore, one change in the metamodel may impact its n different usages. Our 1 : n impact analysis leverages on the power of abstraction of the metamodels [28] to identify those n impacts on the code to co-evolve. We leverage on the knowledge of how the core API code is generated to cover all n different usages of a given metamodel element. Finally, the impact analysis consists in accessing for each impacting metamodel change on an element $e_i$, the set of impacted code parts and their impacted AST nodes.

3.2 Code Co-evolution

Algorithm 2 illustrates the overall co-evolution process. It first retrieves the impacted code parts by a given metamodel change from the mapping table computed by our impact analysis (see Figure 3) that serves to bridge existing change detection approaches with another one.

The change interface considers both atomic and complex changes [19]. Atomic changes are additions, removals, and updates of a metamodel element. Complex changes consist in a sequence of atomic changes combined together. For example, push property is a complex change where a property is moved from a superclass to its subclasses. This is composed of two atomic changes: delete a property and add a property.

With the detected metamodel changes, we run an impact analysis [22] on the code. To this end, we parse the code to access the Abstract Syntax Tree (AST). We then establish a mapping between the metamodel elements $e_1, \ldots, e_n$ and the code AST nodes corresponding to $e_1, \ldots, e_n$, as illustrated in Algorithm 1.

In contrast to the co-evolution of metamodel and models(transformations/constraints) that considers a 1 : 1 relationship between the metamodel elements and their use in the impacted artifacts. Our impact analysis considers a 1 : n relationship where a metamodel element is used in n different forms at the code level.

Indeed, take for example the class Property in the OCL Pivot metamodel. At the code level, an interface Property and a class PropertyImpl are generated. But also in the classes PivotFactoryImpl and PivotPackageImpl, respectively the methods createProperty() and getProperty() are generated. All this n code correspondences of the class Property are identified in Algorithm 1 (Lines 5-11). Other metamodel elements also have n code correspondences, such as attributes. For example, for the attribute implicit in the class Property, a getter and a setter getImplicit()/setImplicit() are generated, but also the method getImplicit() is generated in the class PivotPackageImpl (handled in Algorithm 1 (Lines 13-20)).

Therefore, one change in the metamodel may impact its n different usages. Our 1 : n impact analysis leverages on the power of abstraction of the metamodels [28] to identify those n impacts on the code to co-evolve. We leverage on the knowledge of how the core API code is generated to cover all n different usages of a given metamodel element. Finally, the impact analysis consists in accessing for each impacting metamodel change on an element $e_i$, the set of impacted code parts and their impacted AST nodes.
Algorithm 2 Code co-evolution

1: function Code_Co-Evolution(Set<changes> changes, 
   Map<MME, List<CE>> MT, ResolutionCatalog RC)
2:   for all c ∈ changes do
3:     List<CE> codeImpacts ← MT.get(c.element)
4:     for all cImpact ∈ codeImpacts do
5:         resolutions ← RC.getResolutions(cImpact, c)
6:         if resolutions.size > 1 then
7:             res ← UserDecision(resolutions)
8:         else
9:             if resolutions.size = 1 then
10:                res ← resolutions.first()
11:         end if
12:     end if
13:     cImpact.apply(resolution)
14: end for
15: end for
16: end function

analysis (line 3). Then for each impact, it proposes the appropriate resolutions that can propagate the impacting change (line 5). A developer can choose among the alternative resolutions which one fits her needs. This acts as a user acceptance of the resolution to be applied (line 7). Finally, the chosen resolution is applied on the impacted code part (line 13).

Table 1 depicts our resolutions that propagate to the code metamodel changes that are known to be impacting [22]. The resolutions are inspired from the co-evolution of other artifacts, such as models, constraints, or transformations [2, 5, 6, 11–13, 15, 18, 23–25, 27, 32–34, 49], where they showed to be efficient and useful. Nonetheless, they are further adapted to fit the code co-evolution.

For example, resolutions CR15, CR16, CR17 aims to co-evolve the different code impacts of changing a property type in the metamodel. To react to delete changes in the metamodels, we propose to delete different granularities of the impacted code, such as deleting the direct element, its direct parent expression (whole call path), or its instruction. Here we make sure that no error is introduced by further propagating the resolutions when necessary (e.g., to update method calls due to delete of a parameter) and by adding a cast (e.g., to the expression using a deleted property).

Note that, similarly as for code repair approaches [39], we do not claim completeness of our catalog of resolutions. It rather represents meaningful propagations of the metamodel changes at the code level. Enriching our catalog with other resolutions is left for future work. However, as we will show in section 4, the catalog of resolutions in Table 1 is sufficient and useful in practice, and allows to correctly co-evolve impacted code w.r.t. the developers intent.

3.3 Prototype Implementation

Our approach’s Java implementation handles Ecore/EMF metamodels and Java code. It interfaces with our previous work of change detection [26] to then run an impact analysis on the code. For each impacted part, we propose resolutions. The resolutions are implemented in Java. Both the impact analysis and the resolutions work on the JDT API\textsuperscript{2} to parse and to manipulate the code, application of resolutions.

4 EVALUATION

This section presents the evaluation results of our co-evolution approach. First, we present the dataset and evaluation process. Then, we present the research questions and discuss the obtained results. Time performance of the co-evolution is measured as well.

4.1 Data Set

This section presents the used data set in our evaluation. Our data set covers three case studies covering three different language implementations in Eclipse, namely OCL [37], Modisco [36], and Papyrus [38].

These languages have been developed for more than 10 years and have been evolved several times. OCL is a standard language defined by the Object Management Group (OMG) to specify First-order logic constraints. Modisco is an academic initiative to support development of model-driven tools, verification, and transformation of existing software systems. Papyrus is an industrial project led by CEA\textsuperscript{3} to support model-based simulation, formal testing, safety analysis, etc. Thus, our data set covers standard, academic, and industrial languages. We considered original and evolved versions of the case studies, as shown in Table 2.

The three case studies cover code co-evolution in response to five evolved metamodels respectively, two in the OCL language, one in the Modisco Language, and two in the Papyrus language, as shown in Table 2. The total of applied metamodel changes was 477 in the five metamodels.

For those three case studies, we collected 15 Java projects that were impacted by those five evolving metamodels (i.e., projects that are dependent on the metamodels’ generated core API). We collected the original and evolved Java code of those projects. Three projects were impacted by two metamodels. Thus, in total we considered 18 projects’ co-evolutions due to the five evolved metamodels.

All evolutions in our case studies of the metamodels and code have been performed manually by developers between the different releases that we collected. Thus, we considered the manual code co-evolution as the reference (i.e., ground truth) to which we compare our performed co-evolution.

Table 2 gives details about the selected case studies in particular about their metamodels and the applied changes during evolution. Table 3 further gives details on the size of the projects and code of the original versions that we co-evolve.

\textsuperscript{2}Eclipse Java development tools (JDT): https://www.eclipse.org/jdt/core/

\textsuperscript{3}http://www-list.cea.fr/en/
We evaluate our change propagation-based co-evolution of code by measuring the correctness of our co-evolution. We compare for the same set of code how it was manually co-evolved by developers against our proposed co-evolution. In this experiment, we measure the correctness of our approach by using the two metrics precision and recall that vary from 0 to 1, i.e., 0% to 100%. They are defined as follows:

### 4.2 Evaluation Process

We evaluate our change propagation-based co-evolution of code by measuring the correctness of our co-evolution. We compare for the same set of code how it was manually co-evolved by developers against our proposed co-evolution. In this experiment, we measure the correctness of our approach by using the two metrics precision and recall that vary from 0 to 1, i.e., 0% to 100%. They are defined as follows:

#### Table 1: Catalog of resolutions that propagates the metamodel changes.

<table>
<thead>
<tr>
<th>Impacting Metamodel Changes</th>
<th>Proposed Code Resolutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Delete property p</td>
<td>□ CR1 Remove the direct use of p (e.g., label = s.name + s.m1().p.m2()) → label = s.name + ( (Type_Of_P) s.m1() .m2())</td>
</tr>
<tr>
<td>□ CR2 Remove the statement using p (i.e., if, loop, assignment, etc.)</td>
<td></td>
</tr>
<tr>
<td>□ CR3 Remove the whole call path of p (e.g., label = s.name + s.m1().m2(); p → label = s.name)</td>
<td></td>
</tr>
<tr>
<td>□ CR4 Replace the whole call path of p with a default value (e.g., id = id + s.m1().m2(); p → id = s.id + 0)</td>
<td></td>
</tr>
<tr>
<td>□ CR5 Replace p with another property p' (can be given as input by the developer)</td>
<td></td>
</tr>
<tr>
<td>□ CR6 Rename e in the code</td>
<td></td>
</tr>
<tr>
<td>□ CR7 Introduce a for loop statement to iterate on a collection (e.g., bs.p.value → for(v in bs.p) {v.val})</td>
<td></td>
</tr>
<tr>
<td>□ CR8 Retrieve the first value of a collection (e.g., value = bs.p → value = bs.p.toArray()[0])</td>
<td></td>
</tr>
<tr>
<td>□ CR9 Extend navigation path of p1 (e.g., bs.p1 → bs.ref.p1)</td>
<td></td>
</tr>
<tr>
<td>□ CR10 Reduce navigation path of p1 (e.g., bs.ref.p1 → bs.p1)</td>
<td></td>
</tr>
<tr>
<td>□ CR11 Extend navigation path of p1 and add a for loop (e.g., bs.p1 → for(v in bs.ref) {v.p1})</td>
<td></td>
</tr>
<tr>
<td>□ CR12 Introduce a type test with an If statement (e.g., t.name = s.p.name → if s.p.typeof(Sub) → t.name = (Sub s).p.name)</td>
<td></td>
</tr>
<tr>
<td>□ CR13 Cast p to one specific sub class Sub (e.g., t.name = s.p.name → t.name = (Sub s).p.name)</td>
<td></td>
</tr>
<tr>
<td>□ CR14 Replace p with another property p'</td>
<td></td>
</tr>
<tr>
<td>□ CR15 Change the class type from S to T through ref (e.g., List&lt;S&gt; l = ... → List&lt;T&gt; l = ... )</td>
<td></td>
</tr>
<tr>
<td>□ CR16 Add a cast of p when used as argument of a method (e.g., mCall(p) → mCall( (ExpectedType) p))</td>
<td></td>
</tr>
<tr>
<td>□ CR17 Cast the value for the setter of p to T (e.g., s.setP(v); → s.setP((T) v));</td>
<td></td>
</tr>
</tbody>
</table>

#### Table 2: Details of the metamodels and their evolutions.

<table>
<thead>
<tr>
<th>Case study</th>
<th>Evolved metamodels</th>
<th>Versions</th>
<th>Atomic changes in the metamodel</th>
<th>Complex changes in the metamodel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OCL</strong></td>
<td>Pivot.ecore in project</td>
<td>3.2.2 to 3.4.4</td>
<td>Delegates: 2 classes, 16 properties, 6 super types</td>
<td>1 pull property</td>
</tr>
<tr>
<td></td>
<td>ocl.examples.pivot</td>
<td>3.2.2 to 3.4.4</td>
<td>Renames: 1 class, 5 properties</td>
<td>2 push properties</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PROPERTY changes: 4 types, 2 multiplicities</td>
<td>Total = 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adds: 25 classes, 121 properties, 36 super types</td>
<td>Total = 218</td>
</tr>
<tr>
<td><strong>Modisco</strong></td>
<td>Benchmark.ecore in project</td>
<td>0.9.0 to 0.13.0</td>
<td>Delegates: 6 classes, 19 properties, 5 super types</td>
<td>4 moves property</td>
</tr>
<tr>
<td></td>
<td>modisco.infra.discovery.benchmark</td>
<td>0.9.0 to 0.13.0</td>
<td>Renames: 5 properties</td>
<td>6 pull property</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PROPERTY changes: 2 type</td>
<td>1 push property</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adds: 7 classes, 24 properties, 4 super types</td>
<td>2 extract class</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total = 70 changes</td>
<td>2 extract super class</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total = 12</td>
<td>Total = 12</td>
</tr>
<tr>
<td><strong>Papyrus</strong></td>
<td>ExtendedTypes.ecore in project</td>
<td>0.9.0 to 1.1.0</td>
<td>Delegates: 10 properties, 2 super types</td>
<td>2 pull property</td>
</tr>
<tr>
<td></td>
<td>papyrus.infra.extendedtypes</td>
<td>0.9.0 to 1.1.0</td>
<td>Renames: 3 classes, 2 properties</td>
<td>1 push property</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PROPERTY changes: 2 type</td>
<td>1 extract super class</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Adds: 8 classes, 9 properties, 8 super types</td>
<td>Total = 42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total = 17</td>
<td>Total = 3</td>
</tr>
<tr>
<td></td>
<td>Configuration.ecore in project</td>
<td>0.9.0 to 1.1.0</td>
<td>Delegates: 6 classes, 7 properties, 4 super types</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>papyrus.infra.queries.core.configuration</td>
<td>0.9.0 to 1.1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...
This aims to investigate when our co-evolution diverges from when it fails to propose the expected one by developers? We run an impact analysis to identify the impacted parts of the code. To what extent does our approach correctly co-evolve the impacted code? This measures its correctness level with the precision and recall metrics, i.e., to which extent it proposes the expected resolutions from developers. It also allows us to assess its usefulness. For alternative resolutions, we select the one as close as possible to the expected resolution. This shows the applicability of our co-evolution approach. One to five resolutions were proposed for each impacted part of the code that will be co-evolved.

### 4.3 Research Questions

This section defines two research questions (RQs) to assess in particular the applicability, correctness, and the usefulness of our work. The research questions are as follows:

- **RQ1**: To what extent and how fast can our co-evolution approach handle the impacted parts of the code? If we cannot propose a co-evolution for all impacted parts due to evolving metamodels, we can conclude that our change propagation-based co-evolution of code is insufficient. This assesses the overall applicability of our co-evolution approach.

- **RQ2**: To what extent does our approach correctly co-evolve the impacted code? This measures its correctness level with the precision and recall metrics, i.e., to which extent it proposes the expected resolutions from developers. It also allows us to assess its usefulness. For alternative resolutions, we select the one as close as possible to the expected resolution to be able to compare them afterward.

- **RQ3**: Can our approach still propose relevant co-evolution when it fails to propose the expected one by developers? This aims to investigate, when our co-evolution diverges from the expected one, whether our proposed resolutions are still useful to developers.

### 4.4 Results

We now discuss the results w.r.t. our research questions.

#### 4.4.1 RQ1

With our impact analysis, we were able to find the impacted code parts by the 477 metamodel changes for which resolutions were proposed by our co-evolution approach. One to five resolutions were proposed for each impact in the code. To co-evolve all impacted parts, a total of 983 resolutions were applied during the change propagation. This shows the applicability of our co-evolution approach that was able to handle all different impacts in the code caused by the metamodel evolution changes. Table 4 gives the list of the applied resolutions for each co-evolved project during co-evolution. In total, the 972 applied resolutions were covered by 11 resolution types from our catalog in Table 1, namely CR1, CR2, CR3, CR5, CR6, CR9, CR11, CR12, CR15, CR16, CR17. These resolutions allowed...
The evaluation was run on a Windows 7 PC with a Core i7 3.4GHz and 16GB RAM.

Table 4: Number of applied resolutions in our code co-evolution for each project and per evolved metamodel.

<table>
<thead>
<tr>
<th>Evolved metamodels</th>
<th>Co-evolved projects</th>
<th>N° of applied resolutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCL Papyrus.ecore</td>
<td>P1</td>
<td>CR1:3, CR2:57, CR3:298,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR5:19, CR6:108, CR12:18,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR15:15, CR16:3, CR17:11</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>CR2:10, CR3:8, CR5:2,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR6:5, CR12:2, CR15:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR16:1, CR17:1</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>CR1:3, CR2:91, CR3:5,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR5:18, CR6:1, CR9:10,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR11:1, CR12:1, CR15:7,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR16:3, CR17:1</td>
</tr>
<tr>
<td>OCL Base.ecore</td>
<td>P4</td>
<td>CR2:10, CR5:6</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>CR1:1, CR2:25, CR5:17,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CR9:1, CR11:1, CR12:1</td>
</tr>
<tr>
<td></td>
<td>P6</td>
<td>CR2:9, CR3:1</td>
</tr>
<tr>
<td></td>
<td>P7</td>
<td>CR1:1, CR2:5</td>
</tr>
<tr>
<td></td>
<td>P8</td>
<td>CR5:2</td>
</tr>
<tr>
<td>Modisco Benchmark.</td>
<td>P9</td>
<td>n.a</td>
</tr>
<tr>
<td>.ecore</td>
<td>P10</td>
<td>CR2:13</td>
</tr>
<tr>
<td></td>
<td>P11</td>
<td>CR2:24</td>
</tr>
<tr>
<td>Papyrus ExtendedTypes.</td>
<td>P13</td>
<td>CR2:1, CR6:47</td>
</tr>
<tr>
<td>.ecore</td>
<td>P14</td>
<td>CR6:16</td>
</tr>
<tr>
<td></td>
<td>P15</td>
<td>CR6:14</td>
</tr>
<tr>
<td>Papyrus Configuration.</td>
<td>P16</td>
<td>n.a</td>
</tr>
<tr>
<td>.ecore</td>
<td>P17</td>
<td>CR2:12</td>
</tr>
<tr>
<td></td>
<td>P18</td>
<td>CR2:18</td>
</tr>
</tbody>
</table>

The metamodel changes present in our case studies, their resolutions were never applied, such as CR7, CR8, CR14.

The parsing of the project’s code took from couple of seconds to a minute for the largest project. For each impacting metamodel change, the detection of the impacted code parts took from milliseconds to few seconds, and the application of each resolution was a matter of less than half a second. The evaluation was run on a Windows 7 PC with a Core i7 3.4GHz and 16GB RAM.

4.4.2 RQ2. The metamodel changes present in our case studies (see Table 2 columns 5 and 6) covered different types of changes and evolution scenarios. In particular the impacting changes for which we proposed resolutions in Table 1. To assess to what extent can our approach correctly co-evolve the impacted code w.r.t. the manual co-evolution of the code that developers went through, we measured precision and recall.

Figure 5 depict the reached precision and recall in our 18 scenarios of metamodel evolution and code co-evolution. The projects [P9] and [P16] were not impacted by any metamodel evolution with no performed co-evolution. Thus, we did not measure precision and recall for them. For the rest of the projects, both the measured precision and recall varied from 48% to 100%. Our code co-evolution was able to reach respectively the average of 94.5% and 95.5% of precision and recall. However, as the size of the projects and the metamodels’ evolution varied from small to large, we computed a weighted average to better characterize precision and recall. We computed the weighted average by the number of the metamodel changes for each project, resulting in a weighted average of 87.4% and 88.9% of precision and recall. This means that our applied resolutions propagating the impacting metamodel changes covered the expected resolutions by the developers in 88.9% and were correct in 87.4%. This shows that the catalog of resolutions in Table 1 showed to be both sufficient and useful in co-evolving code in our case studies while meeting the developer’s co-evolution needs.

4.4.3 RQ3. We further looked into the cases where our proposed resolutions mismatched the expected resolutions, i.e., the cause of lowering precision and recall. We observed that for all those impacts in the code caused by non-delete changes in the metamodels, the expected resolutions applied in the co-evolution by the developers were to delete the impacted code. Rather than to delete them, our co-evolution approach was able to successfully maintain them.

For example, in the project [P12], we observed that 28 impacted parts of the code were due to move and rename changes in the metamodel. Listing 6 shows an excerpt of two impacted part of the code from the project [P12] due to: 1) rename discoveryError to discoveryErrors, and 2) our co-evolution approach was able to maintain impacted parts of the code, whereas, the developers’ manual co-evolution consisted in deleting them as the whole project was deleted in the new version. The applied resolutions were replacing the call to discoveryDate to a call from its new class and renaming discoveryError, as shown in Listing 8.
Similarly, in the project [P1], we observed the same cases of 41 deleted code impacted by the metamodel changes, rename, push property, and type change. Listing 7 shows an excerpt of two impacted parts of the code from the project [P1] due to: 1) push property ownedRule and 2) change type of the property specification. The developer manual co-evolutions consisted in deleting the statements where they were used. Instead, we maintained them by changing the type of the variable declaration and by introducing an if statement with a cast, as shown in Listing 9.

The above observations show that developers unnecessarily deleted many impacted parts in the code, while our approach instead successfully maintain them during co-evolution. This is the cause of lowering precision and recall in our case studies, e.g., 48% in the project [P12]. From our point of view, it is surprising to delete code statements, whereas it would have been possible to maintain them (e.g., with rename, move, or push resolutions). This obviously hints on the lack of co-evolution support. Otherwise, they would have been easily maintained in the new version. Another explanation is that the code functionality (or its requirement) became unnecessary in the new version, which explains its deletion.

Whereas these cases lower precision and recall, we do not consider it as wrong co-evolution. It could actually be preferred over simply deleting the impacted code. Indeed, it is more relevant to propagate a rename of a method than to delete its call elsewhere in the code, which our approach was able to achieve for those cases. Nonetheless, it could be possible that for some case simply deleting the impacted code (by non-delete metamodel changes) is desired. In future work, we could propose delete resolutions as an alternative for any impacted code. That is why it is essential for developers to guide the co-evolution process and decide which resolutions best fit their needs.

Listing 6: Excerpt of the impacted code in the project P12.

```
1 // in the class CDOProjectDiscoveryImpl
2 if (baseClass == Discovery.class) {
3     switch (baseFeatureID) {
4         case BenchmarkPackage.DISCOVERY_ITERATION_DISCOVERY_DATE:
5             return ...;
6         case BenchmarkPackage.DISCOVERY_DISCOVERY_ERRORS:
7             return ...;
8         default: return 1;
9     }
10 }
```

Listing 7: Excerpt of the impacted code in the project P1.

```
1 // in the class InvocationBehavior
2 ...
3 (valueSpecification instanceof ExpressionInOCL) {
4     if (valueSpecification instanceof ExpressionInOCL) {
5         ...
6     }
7     if (namedElement instanceof Namespace) {
8         for (Constraint constraint : (Namespace
9             namedElement).getOwnedRule()) {
10             ...
11         }
12     }
13 }
```

5 THREATS TO VALIDITY

This section discusses threats to validity [51].

5.1 Internal Validity

To measure precision and recall, we had to analyze the developers manual co-evolution to identify the expected resolutions. To reduce the risk of misidentifying an expected resolution, for each impacted part of the code, we investigated the whole co-evolved class. In case we did not find it, we further searched in other classes in case the original impacted part of the code was moved into another class. Thus, we aimed at reducing the risk of missing any correspondence between the original and evolved impacted parts of the code. Moreover, as our co-evolution relies on the quality of detected metamodel changes. We analyzed each detected change and checked whether it occurred between the original and evolved metamodels to alleviate the risk of relying on an incorrect metamodel change.

Moreover, in the evaluation, when alternative resolutions were proposed during co-evolution, we selected the one as close as possible to the expected resolution to be able to compare them afterward. Our goal in the evaluation was to assess (RQ1) whether we can co-evolve the impacted code and (RQ2) usefulness of our approach in terms of how far can it allow a developer to reach the same manual co-evolution of our case studies. Assessing how different developers would co-evolve the same impacts, possibly diverging with different resolutions, was not the goal of our evaluation. It is left for a future user study where we would also investigate what influences developers to select a given resolution over another
one. Still, We observed a start of an answer in (RQ3) where alternative resolutions were actually relevant.

5.2 External Validity

We implemented and evaluated our approach for EMF/Ecore metamodels and Java code. Other languages, such as C# or C++, use different syntax but conceptually use the same constructions as in Java. Although we think that the co-evolution would be applicable for other languages, we cannot generalize our results. Further experimentation on other languages is necessary. However, the only requirement to apply our approach to other languages is to have access to the ASTs of the parsed code and to adapt our resolutions to the new ASTs' structure.

Furthermore, our evaluation was performed on SLs and their toolings in the Eclipse platform. Thus, we cannot generalize our findings to all SLs or DSLs neither domain models for instance. However, our approach could be used to co-evolve Java code built on a generated API from a metamodel or a domain model. This is the case for example for tools such as JHipster that generates a complete and modern Web app or microservice architecture based on a domain model and a set of architectural choices. Further evaluations remain necessary here.

5.3 Conclusion Validity

Our evaluation gave promising results, showing that our code co-evolution is fast and useful with a weighted average precision and recall respectively of 87.4% and 88.9%. The evaluation results also showed the usefulness of our approach and the proposed resolutions in our catalog in Table 1. Even though we evaluated it on 18 scenarios of metamodel evolution and code co-evolution. To have more insights and statistical evidence, further evaluation is needed on more case studies.

6 RELATED WORK

The main idea of change propagation was investigated for different purposes, such as by Cubranic et al. [7] who used it to recommend relevant software development artifacts, or Demuth el al. [9] who used it for models co-evolution. In our work we use change propagation to co-evolve code.

In this section, we present the main related work w.r.t. co-evolution. Many approaches proposed to co-evolve models [5, 11, 16, 18, 24, 25, 44, 49], constraints [2, 6, 27, 31, 33], and model transformations [12, 13, 23, 32, 34]. Our work first distinguishes by co-evolving code. It further distinguishes with related work by performing a 1 : n impact analysis. Whereas, a 1 : 1 impact analysis is performed in the approaches of models/constraints(transformations) co-evolution [2, 5, 6, 11–13, 18, 23–25, 27, 32–34, 49].

In this paper, we focus so far the co-evolution on the direction metamodel to code, which is not trivial. Yu et al. [52] proposes to co-evolve the metamodels and the generated API in both directions. However, they do not co-evolve the additional code on top of it, which our approach does.

Existing approaches of code migration are related to our work. We focus on the main existing approaches. Henkel et al. [17] proposed an approach that captures refactoring actions and replay them on the code to migrate. However, they support only the changes renames, moves, and type changes.

Nguyen et al. [40] also proposed an approach that guides developers in adapting code by learning adaptation patterns from previously migrated code. Similarly, Dagenais et al. [8] also uses a recommendation mechanism of code changes by mining them from previously migrated code. Anderson et al. [1] proposed to migrate drivers in response to evolutions in Linux internal libraries. It identifies common changes made in a set of files to extract a generic patch that can be reused on other code parts.

Our current work distinguishes from these code migration approaches [1, 8, 17, 40] by considering and reasoning on the changes at the metamodel level and not at the code level. Thus, our approach treats way less changes to identify the impacted parts in the code than code migration approaches [1, 8, 17, 40]. This is possible thanks to the abstraction offered by the metamodels. Whereas our work considers one change of a given element, n changes must be considered in order to fully migrate the code. To the best of our knowledge, our work is the first attempt to address the challenge of code co-evolution with evolving metamodels in SLs.

7 CONCLUSION

This paper proposed a code co-evolution approach when metamodels evolve. It supports a change propagation-based co-evolution by leverages on the metamodel changes and propagates them on the code. Our code co-evolution was evaluated on 15 projects and five metamodels from three Eclipse language implementations and it was stress tested on 18 scenarios of metamodel evolution and code co-evolution. It showed to be efficient and useful in co-evolving code with a weighted average precision and recall respectively of 87.4% and 88.9%.

As future work, we first plan to enrich our catalogue of resolutions. Although, we envision and present our resolutions similarly as code quick fixes or repair for developers to choose from, we plan to explore ranking heuristics to ease the task of selecting resolutions for developers. Building on that, we can then propose a full automation based on the highest ranked resolutions to compute co-evolutions plans. Finally, we will evaluate our approach on more case studies other than Eclipse language implementations.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the CNRS PEPS, and from the AIS Rennes Météropole under grant no. 190270.
REFERENCES


