

Perspectives of semantic modeling in categories

William Steingartner

Technical University of Košice Faculty of Electrical Engineering and Informatics

* * *

14 January – 16 January 2022

Categories

 $\label{eq:category} \begin{array}{l} \mathcal{C} = (\mathcal{C}_{obj}, \mathcal{C}_{morp}) \text{ is a} \\ \text{structure consisting of} \end{array}$

- class of *objects* \mathscr{C}_{obj} ,
- class of morphisms Cmorp,
- composition binary operation
 o defined on morphisms.

Functor $F: \mathscr{C} \to \mathscr{D}$ is a structure-preserving morphism between categories

$$\begin{array}{l} F_0: \mathscr{C}_{obj} \to \mathscr{D}_{obj} \\ F_1: \mathscr{C}_{morp} \to \mathscr{D}_{morp} \end{array}$$



Categories

- mathematical structures consisting of objects and morphisms between them,
- objects can be various mathematical structures, data structures, types,
- categories have become useful for modeling computations, processes, programs, program systems,
- are basic structures for coalgebraic behavioral models.

Categories in teaching

- quite simple mathematical structures,
- graphical representations useful for illustration of examples.
- understandable for our students.

Э



https://www.nasa.gov/consortium/CategoryTheory

E

Formal semantics

- provides unambiguous meaning of programs written in programming language,
- helps designers to prepare good and useful programming languages,
- serves for designers to design correct compilers,
- encourages users/programmers how to use language constructions properly.

Semantic methods

- denotational semantics,
- operational semantics,
- natural semantics,
- axiomatic semantics,
- action semantics,
- game semantics,
- . . .

Categorical semantics

- denotational semantics uses category of types where objects are types and morphisms are functions,
- algebraic semantics uses institutions as complex structures based on categories of signatures,
- game semantics uses category of arenas.

Basic ideas of our approach

Why categorical semantics

- provides illustrative view of dynamics of states,
- provides simply understandable mathematical model of programs,
- appropriate for designers of compilers,
- serves for creating skills to work with formal methods.

Construction of category of states

- we consider simple imperative language,
- our language has only two implicit types,
- for now, we do not consider exception, jumps and recursion,
- we construct category of states,
- environment of procedures is constructed as category of categories,
- so simplified model is understandable without losing exactness.

Categorical denotational semantics of imperative languages

Categorical representation

- formulation of meaning indicates (determines) a construction of a categorical model for a given program,
- categorical model consists of:
 - objects states during the program execution,
 - morphisms, which express the relations between objects steps of computations.

The language Jane

$n \in \mathbf{Num}$	$x \in \mathbf{Var}$
$e \in \mathbf{Expr}$	$b \in \mathbf{Bexpr}$
$S \in \mathbf{Statm}$	$D \in \mathbf{Decl}$

Steingartner, W., Novitzká, V., Bačíková, M., Korečko, Š., New approach to categorical

semantics for procedural languages, Computing and Informatics, 36(6), 2017, pp. 1385-1414,

DOI: 10.4149/cai_2017.6.1385

William Steingartner

Language Jane – Syntax

The elements $n \in \mathbf{Num}$ and $x \in \mathbf{Var}$ have no internal structure from semantic point of view.

The syntactic domain \mathbf{Expr} consists of all well-formed arithmetic expressions created by the following production rule

e ::= n | x | e + e | e - e | e * e.

Boolean expression from **Bexpr** can be of the following structure:

```
b ::= \texttt{false} \mid \texttt{true} \mid e = e \mid e \leq e \mid \neg b \mid b \land b.
```

The variables used in programs have to be declared. We consider $D \in \mathbf{Decl}$ as a sequence of declarations:

```
D ::= \operatorname{var} x; D \mid \varepsilon.
```

As the statements $S \in$ **Statm** we consider five Dijkstra's statements together with a block statement and an input statement:

S ::= x := e |skip|S; S| if b then S else S| while b do S| begin D; S end| input x.

Categorical denotational semantics of imperative languages

Categorical model

- we construct operational model of Jane as the category \mathscr{C}_{State} of states,
- we assign to states their representation,
- because of block structure of *Jane*, we have to consider also a level of block nesting $(l \in \mathbf{Level}, \mathbf{Level} \subseteq \mathbf{N})$,
- representation of type *State* has to express variable, its value with respect to the actual nesting level.

Specification of states

State

- can be considered as some abstraction of computer memory.
- change of state means change of a value in memory,
- because of block structure of Jane, we have to consider also a level of block nesting,
- every variable occurring in a program has to be allocated,
- we can assign and modify a value of allocated variable inducing change of state.

The signature Σ_{State} for states $\Sigma_{State} =$ types: State, Var, Value $opns: init: \rightarrow State$ $alloc: Var, State \rightarrow State$ $qet: Var, State \rightarrow Value$ $del: State \rightarrow State$

ъ

Categorical denotational semantics of imperative languages

States and their representation

- the state expresses an abstraction of memory: each step of program execution is characterized by the current state,
- the nesting level in the state allows us to create an environment of variables and distinguish locally declared variables from global ones,
- each state s is an element of the semantic domain $s \in$ **State** and it is represented as a function

$s: \mathbf{Var} \times \mathbf{Level} \to \mathbf{Value}$

$$= \langle ((x_1, 1), v_1), \dots, ((x_n, l), v_n) \rangle$$



э.

s

Representation of operations

The operation [*init*]

 $\llbracket init \rrbracket = s_0 = \langle ((\bot, 1), \bot) \rangle$

creates the initial state of a program with no declared variable.



The operation [alloc]

 $[alloc](x,s) = s \diamond ((x,l), \bot),$

sets actual nesting level to declared variable. Because of undefined value of declared variable, the operation [alloc] does not change the state.



з.

Representation of operations

The operation [get] returns a value of a variable declared on the highest nesting level,

 $\llbracket get \rrbracket (x, \langle \dots, ((x, l_i), v_i), \dots, ((x, l_k), v_k), \dots \rangle) = v_k,$

where $l_i < l_k, i < k$ for all *i*, from the definition of state.

The operation $[\![del]\!]$ deallocates (forgets) all variables declared on the highest nesting level l_j :

 $\llbracket del \rrbracket (s \diamond \langle ((x_i, l_j), v_k), \dots, ((x_n, l_j), v_m) \rangle) = s.$



Declarations

Declarations

A declaration

 $\operatorname{var} x$

is represented as an endomorphism:

$$[\texttt{var } x]\!]_D: s \to s$$

for a given state *s* and defined by

$$\llbracket \texttt{var } x \rrbracket s = \llbracket alloc \rrbracket (x, s).$$

A sequence of declarations:

$$\llbracket \texttt{var} \ x; D \rrbracket s = \llbracket D \rrbracket \circ \llbracket alloc(x, s) \rrbracket.$$

A declaration creates a new entry for declared variable with the actual level of nesting and an undefined value

$$((x,l), \perp)$$
.

Э

Semantics of statements

variable assignment



Semantics of statements

• sequence of statements

$$\llbracket S_1; S_2 \rrbracket s = \begin{cases} s', & \text{if } \llbracket S_1 \rrbracket s = s'' \text{ and } \llbracket S_2 \rrbracket s'' = s', \\ s_{\perp}, & \text{if } \llbracket S_1 \rrbracket s = s_{\perp}, \text{ or} \\ & \text{if } \llbracket S_1 \rrbracket s = s'' \text{ and } \llbracket S_2 \rrbracket s'' = s_{\perp}. \end{cases}$$



Semantics of statements

conditional statement

$$\begin{bmatrix} \text{if } b \text{ then } S_1 \text{ else } S_2 \end{bmatrix} s = \begin{cases} \begin{bmatrix} S_1 \end{bmatrix} s, & \text{if } \begin{bmatrix} b \end{bmatrix} s = \text{true}, \\ \begin{bmatrix} S_2 \end{bmatrix} s, & \text{if } \begin{bmatrix} b \end{bmatrix} s = \text{false}, \\ s_{\perp}, & \text{otherwise.} \end{cases}$$

$$\begin{bmatrix} S_1 \end{bmatrix} \xrightarrow{s}_{1} \xrightarrow{s}_{1} \xrightarrow{s}_{2} \\ \begin{bmatrix} b \end{bmatrix} s = \text{true} \\ (a) & (b) \end{bmatrix}$$

Statements

 $[\![\texttt{while } b \texttt{ do } S]\!]s = [\![\texttt{if } b \texttt{ then } (S,\texttt{while } b \texttt{ do } S) \texttt{ else skip}]\!]$

$$\llbracket \texttt{input } x \rrbracket s = \begin{cases} s' = s [v/x], \\ s_{\perp}, \end{cases}$$

for $((x, \max l), v') \in s$, otherwise.



William Steingartner

Block statement

 $\texttt{begin}\ D,S \texttt{ end}$

The following is a summary of the four steps used to execute of unnamed blocks.

• Nesting level *l* is incremented. We represent this step by fictive entry in state table

 $\left(\left(\texttt{begin}, l+1\right), \bot\right)$

i.e. endomorphism $State \rightarrow State$.

- Local declarations are elaborated on nesting level l + 1.
- The body S of block is performed.
- Locally declared variables are forgotten at the end of block. We model this situation using operation [[del]].

The semantics:

 $[\![\operatorname{begin}\,D,S\,\operatorname{end}\,]\!]s=[\![\operatorname{del}\,]\!]\circ[\![\,S\,]\!]\circ[\![\,D\,]\!](s\diamond\langle((\operatorname{begin},l+1),\bot)\rangle)$

Constructing the category

Now we can define the category \mathscr{C}_{State} of states as follows:

- category objects are states as sequences of tuples for variables together with special state s_{\perp} ,
- category morphisms are functions $[S]: s \to s'$.

The category \mathscr{C}_{State} has the following properties:

- the special object $s_{\perp} = \langle ((\perp, \perp), \perp) \rangle$, an undefined state, is a terminal object of our category, from any object there is a unique morphism to this state,
- the initial state $s_0 = \langle ((\bot, 1), \bot) \rangle$ is the initial object of our category,
- the category \mathscr{C}_{State} has no products, because a program written in Jane cannot be simultaneously in more than one state.

We can state that \mathscr{C}_{State} is a category without products and with initial and terminal objects.

э.

Semantics of procedures



ъ.

• • • • • • • • • • • • •

```
var x; var y;
input x;
input y;
if x \le y then
   begin
    z := x;
    x := y;
    y := z;
   end
 else
   skip;
```

We consider values 3 and 5 for variables x and y, resp.

Categorical representation of program



э

States during program execution



E

Categorical operational semantics

Coalgebraic approach

• coalgebras are defined as arrows from the state space (X) to the image of state space in the endofunctor F (determined by the signature):

 $\langle \llbracket sel_1 \rrbracket, \ldots, \llbracket sel_n \rrbracket \rangle : X \to FX$

- we define coalgebras above the base category, whose objects create a state space and whose morphisms are transitions,
- coalgebras provide observable properties and are one of the tools for modeling the behavior of dynamical systems,
- each individual step of program execution is expressed by the application of a polynomial endofunctor in the category of configurations.

Steingartner, W., Novitzká, V., Schreiner W., Coalgebraic Operational Semantics for an Imperative Language, Computing and Informatics, 38(5), 2019, pp. 1181-1209, DOI: 10.31577/cai_2019_5_1181

3

Categorical operational semantics

State space and its representation

- we consider the data type of configurations as the state space,
- memory expresses one moment of program execution:

 $Memory = \{m : Var \times Level \to Value\},\$

representation of state space is a set

 $Config = Program \times Memory \times Input \times Output,$

where configuration is given as follows:

 $\mathit{config} = \left([\![D^*; S^*]\!], m, i^*, o^* \right).$

Semantics of statements

We define the execution of one step by morphism [next]:

 $\llbracket next \rrbracket$: Config \rightarrow Config.

• variable assignment x := e:

$$[next]]([\![x:=e;S^*]\!],m,i^*,o^*) = ([\![S^*]\!],m',i^*,o^*),$$

where

$$m' = \begin{cases} m \left[((x, Highest(m, x)), v) \mapsto ((x, Highest(m, x)), \llbracket e \rrbracket m) \right] \\ & \text{if } Defined(m, x), \\ m_{\perp}, & \text{otherwise,} \end{cases}$$

empty statement

$$[\operatorname{\mathit{next}}]\!]([\![\operatorname{skip};S^*]\!],m,i^*,o^*)=([\![S^*]\!],m,i^*,o^*)$$

э

Semantics of statements

$$next \,]\!]([\![(S_1; S_2); S^*]\!], m, i^*, o^*) = \begin{cases} ([\![S_2; S^*]\!], m', i^{\prime *}, o^{\prime *}), \\ \text{if } \langle S_1, m \rangle \Rightarrow m', \\ ([\![S_1'; S_2; S^*]\!], m', i^{\prime *}, o^{\prime *}), \\ \text{if } \langle S_1, m \rangle \Rightarrow \langle S_1', m' \rangle, \end{cases}$$

conditional statement

 $[\![\operatorname{\mathit{next}}]\!]([\![\operatorname{if}\ b\ {\tt then}\ S_1\ {\tt else}\ S_2;S^*\,]\!],m,i^*,o^*)=$

$$\begin{array}{ll} (\llbracket S_1; S^* \rrbracket, m, i^*, o^*), & \text{if } \llbracket b \rrbracket m = \mathbf{true}, \\ (\llbracket S_2; S^* \rrbracket, m, i^*, o^*), & \text{if } \llbracket b \rrbracket m = \mathbf{false}, \\ m_{\perp}, & \text{otherwise} \end{array}$$

イロン イロン イヨン イヨン ヨー

Semantics of statements

user input – statement read

 $[\![\mathit{read}\,]\!]:\mathbf{Config}\to\mathbf{Config}^{\mathbf{Value}}$

$$[[read]]([[read x; S^*]], m, i^*, o^*) = \begin{cases} \lambda v' \cdot ([[S^*]], m', tail(i^*), o^*) \\ \text{if } Defined(m, x), \\ ([[S^*]], m_{\perp}, tail(i^*), o^*), \\ \text{otherwise,} \end{cases}$$

where $m' = m[((x, \mathit{Highest}(m, x)), v) \mapsto ((x, \mathit{Highest}(m, x)), v')],$

user output - statement print

 $\llbracket print \rrbracket : \mathbf{Config} \to \mathbf{Value} \times \mathbf{Config}$

 $\llbracket \textit{print} \, \rrbracket(\llbracket \texttt{print} \; e; S^* \, \rrbracket, m, i^*, o^*) = (\llbracket e \, \rrbracket m, (\llbracket S^* \, \rrbracket, m, i^*, (\llbracket e \, \rrbracket m; o^*)))$

Coalgebra for language *Jane*

Coalgebra for language Jane

Category of configurations *Config* consists of:

- objects configurations $confiq = (\llbracket D^*; S^* \rrbracket, m, i^*, o^*),$
- arrows morphisms [[next]], [[read]], [[print]] a [[abort]].

Polynomial endofunctor (over the category of configurations):

$$\langle \llbracket abort \rrbracket, \llbracket print \rrbracket, \llbracket next \rrbracket, \llbracket input \rrbracket \rangle : \mathbf{Config} \to Q(\mathbf{Config}),$$

 $Q(\text{Config}) = 1 + \text{Config} + O \times \text{Config} + \text{Config}^{I}$.

 $Q(config) = [abort](config) \quad Q(config) = [next](config)$ $Q(config) = [print](config) \quad Q(config) = [read](config)$

3

Categorical operational semantics – Example

```
var x; var y;
read x;
read y;
                        For simplicity we introduce the following substitutions:
if x \le y then
                                    D_1 = \operatorname{var} x; D_2 = \operatorname{var} y;
    begin
                                    S_1 = \operatorname{read} x; \ S_2 = \operatorname{read} y;
      var z;
                                    S_3 = \text{if } x \le y \text{ then begin var } z;
                                        z := x; x := y; y := z end else skip
      z := x;
                                    S_4 = \text{print } x
      x := y;
      y := z;
                        and we consider values 3 and 5 for variables x and y, resp.
    end
  else
    skip;
print x;
```

The initial configuration is

$$config_0 = (\llbracket D_1; D_2; S_1; S_2; S_3; S_4 \rrbracket, m_0, i^*, o^*).$$

Each application of the endofunctor Q represents one step of program execution. First, the individual declarations and user inputs are processed in separate steps:

$$\begin{split} Q(\operatorname{config}_0) &= \llbracket \operatorname{next} \rrbracket(\operatorname{config}_0) = \operatorname{config}_1 = \\ &= (\llbracket D_2; S_1; S_2; S_3; S_4 \rrbracket, \llbracket \operatorname{var} x \rrbracket m_0, (\mathbf{3}, \mathbf{5}), \varepsilon) \end{split}$$

$$\begin{array}{ll} Q(config_1) &= [\![next]\!](config_1) = config_2 = \\ &= ([\![S_1; S_2; S_3; S_4]\!], [\![var \ y \]\!]m_1, ({\bf 3}, {\bf 5}), \varepsilon), \\ Q(config_2) &= [\![read]\!](config_2) = config_3 = \\ &= ([\![S_2; S_3; S_4]\!], m_3, ({\bf 5}), \varepsilon), \end{array}$$

$$\begin{array}{ll} Q(\operatorname{config}_3) &= [\![\operatorname{read}\,]\!](\operatorname{config}_3) = \operatorname{config}_4 = \\ &= ([\![\,S_3;S_4\,]\!], m_4, \varepsilon, \varepsilon), \end{array}$$

$$\begin{array}{lll} Q(\operatorname{config}_4) &= \llbracket \operatorname{next} \rrbracket(\operatorname{config}_4) = \operatorname{config}_5 = \\ &= (\llbracket \operatorname{begin} \operatorname{var} z; z := x; x := y; y := z \operatorname{end}; S_4 \rrbracket, m_4, \varepsilon, \varepsilon), \\ Q(\operatorname{config}_5) &= \llbracket \operatorname{next} \rrbracket(\operatorname{config}_5) = \operatorname{config}_6 = \\ &= (\llbracket \operatorname{var} z; z := x; x := y; y := z \operatorname{end}; S_4 \rrbracket, m_5, \varepsilon, \varepsilon), \\ Q(\operatorname{config}_6) &= \llbracket \operatorname{next} \rrbracket(\operatorname{config}_6) = \operatorname{config}_7 = \\ &= (\llbracket z := x; x := y; y := z \operatorname{end}; S_4 \rrbracket, m_6, \varepsilon, \varepsilon), \\ Q(\operatorname{config}_7) &= \llbracket \operatorname{next} \rrbracket(\operatorname{config}_7) = \operatorname{config}_8 = \\ &= (\llbracket x := y; y := z \operatorname{end}; S_4 \rrbracket, m_7, \varepsilon, \varepsilon), \\ Q(\operatorname{config}_8) &= \llbracket \operatorname{next} \rrbracket(\operatorname{config}_8) = \operatorname{config}_9 = \\ &= (\llbracket y := z \operatorname{end}; S_4 \rrbracket, m_8, \varepsilon, \varepsilon), \\ Q(\operatorname{config}_9) &= \llbracket \operatorname{next} \rrbracket(\operatorname{config}_9) = \operatorname{config}_{10} = \\ &= (\llbracket \operatorname{end}; S_4 \rrbracket, m_9, \varepsilon, \varepsilon), \\ Q(\operatorname{config}_1) &= \llbracket \operatorname{next} \rrbracket(\operatorname{config}_{10}) = \operatorname{config}_{11} = \\ &= (\llbracket S_4 \rrbracket, \llbracket \operatorname{end} \rrbracket m_9, \varepsilon, \varepsilon). \end{array}$$

Ę



Ę

Modeling of recursive computations

Modeling of recursive computations

- we model recursion development and subsequent calculation using algebras and coalgebras and their properties,
- there is exactly one morphism from the initial algebra to any algebra (A, a) *catamorphism*, in the calculation it acts as an iterator (deconstructor) a function that provides elements of the structure,
- from any coalgebra (U,φ) there exists one unique morphism into the final coalgebra *anamorphism*, in calculations it acts as a coiterator (constructor) a function that creates a structure,
- the composition of catamorphism and anamorphism creates a new morphism – *hylomorphism*, which represents a recursive function – by creating complex data structures and then processing them.

Steingartner, W., Macko, P., Some New Approaches in Functional Programming Using Algebras and Coalgebras, Electronic Notes in Theoretical Computer Science, 279(3), 2011, pp. 41–62

Categorical modeling in differential calculus

- the relationship between functions and their derivatives expresses a categorical model,
- we model functions and their derivatives as objects in the category of morphisms Der[→] over the base category Der,
- the relationship between the two categories is expressed by a codomain functor:

$$Cod: \mathscr{D}er^{\rightarrow} \rightarrow \mathscr{D}er$$

Steingartner, W., Galinec, D., The Rôle of Categorical Structures in Infinitesimal Calculus, Journal of Applied Mathematics and Computational Mechanics, 12(1), 2013, pp. 107–119

William Steingartner

Perspectives of semantic modeling in categories 37/43

Modeling of component systems

- interfaces and interactions between components (algebraic specifications) – interface category (objects are interfaces, morphisms are interactions),
- contracts for component composition and interaction (I) we extend interface specifications with assumptions and guarantees or (II) we express them as formulas in predicate linear logic,
- Operation of the second sec

3	dependencies
2	contracts
1	interfaces

Steingartner, W., Novitzká, V., Benčková, M., Prazňák, P., Considerations and Ideas in

Component Programming – Towards to Formal Specification, 25th CECIIS, 2014, pp. 332–339

William Steingartner

Denotational semantics of concatenative language

Research in the field of concatenative languages

- we designed a simple concatenative / compositional language KKJ,
- the language has a compositional character the syntactic concatenation of programs corresponds to the semantic composition of functions,
- KKJ language syntax:

 $e ::= \varepsilon \mid i \mid n \mid \{e\} \mid e \ e$

- the state of memory is expressed by the stack the program gradually changes the contents of the stack,
- in our research we constructed classical denotational semantics for the KKJ language as the first step in research.

Mihelič, J., Steingartner, W., Novitzká, V., A denotational semantics of a concatenative / compositional programming language, Acta Politechnica Hungarica, 18(4), 2021, pp. 231–250, DOI: 10.12700/APH.18.4.2021.4.13

Application of research into teaching

Software to support the teaching of formal semantics

- we also apply published results focused on categorical semantics and semantic modeling in the teaching process,
- our main goal is to implement and deploy a comprehensive learning environment – an interactive software package that will allow illustrative and understandable use of semantic methods,
- the mentioned software package will provide full-fledged modules for working with individual semantic methods and principles, which are presented in the teaching.

Steingartner, W., Novitzká, V., A survey of teaching tools for the course on the Semantics of Programming Languages, In: Mathematical Modelling in Physics and Engineering, Częstochowa, Poland, Politechnika Częstochowska, pp. 40–47. *Invited lecture*.

Application of research into teaching

Software to support the teaching of formal semantics



э

Conclusion

Achieved results

- categorical models for programming languages: semantic methods for imperative languages,
- semantic modeling of recursive computations, the relationship of functions and their derivatives, the properties of categorical models of components and the relationships between them and denotational semantics for a new concatenative language,
- application of achieved results to teaching.

Areas for future research

- modeling of component systems,
- semantic modeling for concatenative and some domain-specific languages,
- modeling of properties of mathematical objects,
- application of results in the field of construction of reliable programs.

Thank You for Your attention.

E