



**HAL**  
open science

# An Institution-Independent Proof of the Beth Definability Theorem

Marc Aiguier, Fabrice Barbier

► **To cite this version:**

Marc Aiguier, Fabrice Barbier. An Institution-Independent Proof of the Beth Definability Theorem. *Studia Logica*, 2007, 85 (3), pp.333–359. hal-00341973

**HAL Id: hal-00341973**

**<https://hal.science/hal-00341973>**

Submitted on 19 Jul 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An institution-independent Proof of the Beth Definability Theorem

Marc Aiguier and Fabrice Barbier

<sup>1</sup>Université d'Évry, LaMI CNRS UMR 8042,

523 pl. des Terrasses F-91000 Évry

{aiguier,fbarbier}@lami.univ-evry.fr

fax number: (+33) 1 60 87 37 89

June 30, 2006

## Abstract

A few results generalising well-known conventional model theory ones have been obtained in the framework of institutions these last two decades (*e.g.* *Craig interpolation, ultraproduct, elementary diagrams*). In this paper, we propose a generalised institution-independent version of the Beth definability theorem.

**Keywords:** Beth definability, Craig interpolation, institutions, institution morphisms and comorphisms, inclusive category

## 1 Introduction

A few results generalising well-known conventional model theory ones have been obtained in the framework of institutions [15] these last two decades (*e.g.* [8, 18, 21, 16]). In order to continue this generalisation work of important results from conventional model theory, we present in this paper a generalised institution-independent version of the Beth definability theorem.

The theory of institutions extends Barwise's abstract model theory [3] to computer science. This extension is manifold:

- Institutions include both notions of signature (related to the notion of software interface) and signature morphism (to structure softwares).
- Sentences are only defined as members of a set. This means sets of sentences are neither necessarily closed under the classic logical symbols in  $\{\neg, \wedge, \vee, \Rightarrow, \exists, \forall\}$  nor restricted to them. This allows for a larger family of logics to be taken into account such as Horn clause logic or modal logics.

- Finally, institutions only preserve the renaming property extended to any signature morphism and called *satisfaction condition*.

The original goals of institutions were to generalise results both in computer science and model theory. However, it is mainly in computer science that this task has been accomplished. Despite of its importance for specification theory, the problem of generalising conventional model theory results within the framework of institutions have only been tackled by some isolated works. As far as we know, these are the ones:

- Tarlecki's works [20, 21] which generalise in a particular form of institution some classic algebraic results such as the Birkhoff theorem (equivalence between equational theories and varieties), the McKinsley theorem (equivalence between universal Horn theories and quasi-varieties) and the Mal'cev theorem (existence of initial term models in universal Horn theories),
- Salibra and Scollo's works [17, 18] which deal with relationships between Craig-style interpolation, compactness and Löwein-Skolem properties in a relaxed form of institutions called pre-institutions,
- Diaconescu's recent works [6, 7, 8, 9] which throw the basis for a real study of model theory within the framework of institutions.

Among all the conventional model theory results which are of interest for computer science is the Craig interpolation theorem. Restated in order to better suit computer science's needs as a property over sets of formulae and any kind of signature morphisms, it has been shown to be strongly linked to completeness of structured inference systems [4, 5] and to some aspect of modularity (faithfulness) [10, 11, 24, 25]. It has recently been directly proved in the framework of Birkhoff institutions<sup>1</sup> by R. Diaconescu [8]. Unlike a former result by A. Salibra and G. Scollo [18], R. Diaconescu's result doesn't require negation and is thus suitable for logics without it such as equational logic and Horn clause logic.

In conventional model theory, an important consequence of the Craig interpolation theorem is the Beth definability theorem. This theorem provides an answer to the question to what extent implicit definitions can be made explicit. For example, when one wants to formalize a theory, the very first step is to fix the language, that is deciding which notions are primitives, the others having to be defined from them. But how can one check useless symbols haven't been introduced? This is a problem of crucial importance in specification theory and artificial intelligence and the Beth definability theorem is a tool to solve it. Considering the assets of the institutional framework for model theory (abstractness, logic-independence) we study in this paper both definability notions (implicit and explicit) from an institution-independent point of view. We address their generalisation from two different angles.

---

<sup>1</sup>Birkhoff institutions are a particular form of institutions model classes of which are closed under some algebraic operations such as varieties and quasi-varieties are.)

1. First we define both definability notions in the restricted case of inclusion signature morphisms. To this end, we make use of the notion of strongly inclusive category [10] in order to give a categorical definition of both set-theoretical notions of inclusion and difference.
2. Since non-injective morphisms are of great importance in specification theory, we then generalise both definability notions to any kind of signature morphisms.

We also study the preservation of Beth definability theorem through institution morphisms and comorphisms. This allows inheritance of this theorem from one logic to another, both ones being presented as institutions and linked by an institution morphism.

This paper is structured as follows: Section 2 reviews some concepts, notations and terminology about institutions and institution morphisms and comorphisms which are used by this work. Section 3 reviews the institution-independent model theoretic concept of Craig interpolation. Taking inspiration from a similar result on institution transformations [18], we formulate preservation theorems for Craig interpolation property through institution morphisms and comorphisms. Section 4 formulates a general institution-independent version of the Beth definability theorem and proves it as a consequence of the Craig interpolation property. As for Craig interpolation we develop a preservation theorem for Beth definability through institution morphisms.

## 2 Institutions

Intuitively, the theory of institutions abstracts the semantical part of logical systems according to the needs of software specification in which changes of signatures occur frequently. In this section we review and define some of the basic notions on institutions in use in this paper.

### 2.1 Basic definitions and examples

An institution [15] consists of a category of signatures such that associated with each signature are sentences, models and a relationship of satisfaction that, in a certain sense, is invariant under change of signature. More precisely, this means that a change of signature (by a signature morphism) induces “consistent” changes in sentences and models in a sense made precise by the “Satisfaction Condition” in Definition 2.1 below. This goes a step beyond Tarski’s classic “semantic definition of truth” [23] and also generalises Barwise’s “Translation Axiom” [3]. Moreover, it is fundamental that sentences translate in the same direction as the change of notation, whereas models translate in the opposite direction (think of signature enrichment and model reduction). This is the reason for the functor  $Mod$  in Definition 2.1 below to be contravariant. For the sake of generalisation, signatures are simply defined as objects of a category

and sentences built over a signature are simply required to form a set. All other contingencies such as inductive definition of sentences are not considered. Similarly, models are simply seen as objects of a category, *i.e.* no particular structure is imposed on them. Finally, properties satisfied by a given class of models are characterized through a binary relation between models and sentences of a given signature. More formally, an institution is defined as follows:

**Definition 2.1 (Institution)** *An institution  $\mathcal{I} = (Sig, Sen, Mod, \models)$  consists of*

- *a category  $Sig$ , objects of which are called signatures,*
- *a functor  $Sen : Sig \rightarrow Set$  giving for each signature a set, elements of which are called sentences,*
- *a contravariant functor  $Mod : Sig^{op} \rightarrow Cat$  giving for each signature a category, objects and arrows of which are called  $\Sigma$ -models and  $\Sigma$ -morphisms respectively, and*
- *a  $|Sig|$ -indexed family of relations  $\models_{\Sigma} \subseteq |Mod(\Sigma)| \times Sen(\Sigma)$  called satisfaction relation,*

*such that the following property holds:*

$$\forall \sigma : \Sigma \rightarrow \Sigma', \forall \mathcal{M}' \in |Mod(\Sigma')|, \forall \varphi \in Sen(\Sigma),$$

$$\mathcal{M}' \models_{\Sigma'} Sen(\sigma)(\varphi) \Leftrightarrow Mod(\sigma)(\mathcal{M}') \models_{\Sigma} \varphi$$

**Example 2.2** *The following examples of institutions are of particular importance for computer science. Many other examples can be found in the literature (e.g. [15, 22]).*

**Propositional Logic (PL)** *Signatures and signature morphisms are sets of propositional variables and functions between them respectively.*

*Given a signature  $\Sigma$ , the set of  $\Sigma$ -sentences is the least set of sentences finitely built over propositional variables in  $\Sigma$  and Boolean connectives in  $\{\neg, \vee\}$ . Given a signature morphism  $\sigma : \Sigma \rightarrow \Sigma'$ ,  $Sen(\sigma)$  translates  $\Sigma$ -formulae to  $\Sigma'$ -formulae by renaming propositional variables according to  $\sigma$ .*

*Given a signature  $\Sigma$ , the category of  $\Sigma$ -models is the category of mappings  $\nu : \Sigma \rightarrow \{0, 1\}$ <sup>2</sup> with identities as morphisms. Given a signature morphism  $\sigma : \Sigma \rightarrow \Sigma'$ , the forgetful functor  $Mod(\sigma)$  maps a  $\Sigma'$ -model  $\nu'$  to the  $\Sigma$ -model  $\nu = \nu' \circ \sigma$ .*

*Finally, satisfaction is the usual propositional satisfaction.*

**Many-sorted First Order Logic (FOL)** *Signatures are triples  $(S, F, P)$  where  $S$  is a set of sorts, and  $F$  and  $P$  are sets of function and predicate names*

---

<sup>2</sup> $\{0, 1\}$  are the usual truth-values.

respectively, both with arities in  $S^* \times S$  and  $S^+$  respectively.<sup>3</sup> Signature morphisms  $\sigma : (S, F, P) \rightarrow (S', F', P')$  consist of three functions between sets of sorts, sets of functions and sets of predicates respectively, the last two preserving arities.

Given a signature  $\Sigma = (S, F, P)$ , the  $\Sigma$ -atoms are of two possible forms:  $t_1 = t_2$  where  $t_1, t_2 \in T_F(X)_s$ <sup>4</sup> ( $s \in S$ ), and  $p(t_1, \dots, t_n)$  where  $p : s_1 \times \dots \times s_n \in P$  and  $t_i \in T_F(X)_{s_i}$  ( $1 \leq i \leq n, s_i \in S$ ). The set of  $\Sigma$ -sentences is the least set of formulae built over the set of  $\Sigma$ -atoms by finitely applying Boolean connectives in  $\{\neg, \vee\}$  and the quantifier  $\forall$ .

Given a signature  $\Sigma = (S, F, P)$ , a  $\Sigma$ -model  $\mathcal{M}$  is a family  $M = (M_s)_{s \in S}$  of sets (one for every  $s \in S$ ), each one equipped with a function  $f^{\mathcal{M}} : M_{s_1} \times \dots \times M_{s_n} \rightarrow M_s$  for every  $f : s_1 \times \dots \times s_n \rightarrow s \in F$  and with a  $n$ -ary relation  $p^{\mathcal{M}} \subseteq M_{s_1} \times \dots \times M_{s_n}$  for every  $p : s_1 \times \dots \times s_n \in P$ . Given a signature morphism  $\sigma : \Sigma = (S, F, P) \rightarrow \Sigma' = (S', F', P')$  and a  $\Sigma'$ -model  $\mathcal{M}'$ ,  $\text{Mod}(\sigma)(\mathcal{M}')$  is the  $\Sigma$ -model  $\mathcal{M}$  defined for every  $s \in S$  by  $M_s = M'_s$ , and for every function name  $f \in F$  and predicate name  $p \in P$ , by  $f^{\mathcal{M}} = \sigma(f)^{\mathcal{M}'}$  and  $p^{\mathcal{M}} = \sigma(p)^{\mathcal{M}'}$ .

Finally, satisfaction is the usual first-order satisfaction.

**Horn Clause Logic (HCL)** An universal Horn sentence for a signature  $\Sigma$  in **FOOL** is a  $\Sigma$ -sentence of the form  $\Gamma \Rightarrow \alpha$  where  $\Gamma$  is a finite conjunction of  $\Sigma$ -atoms and  $\alpha$  is a  $\Sigma$ -atom. The institution of Horn clause logic is the sub-institution of **FOOL** whose signatures and models are those of **FOOL** and sentences are restricted to the universal Horn sentences.

**Equational Logic (EQL)** An algebraic signature  $(S, F)$  simply is a **FOOL** signature without predicate symbols. The institution of equational logic is the sub-institution of **FOOL** whose signatures and models are algebraic signatures and algebras respectively, and sentences are restricted to equations.

**Rewriting Logic (RWL)** Given an algebraic signature  $\Sigma = (S, F)$ ,  $\Sigma$ -sentences are formulae of the form  $\varphi : t_1 \rightarrow t'_1 \wedge \dots \wedge t_n \rightarrow t'_n \Rightarrow t \rightarrow t'$  where  $t_i, t'_i \in T_F(X)_{s_i}$  ( $1 \leq i \leq n, s_i \in S$ ) and  $t, t' \in T_F(X)_s$  ( $s \in S$ ). Models of rewriting logic are preorder models, i.e. given a signature  $\Sigma = (S, F)$ ,  $\text{Mod}(\Sigma)$  is the category of  $\Sigma$ -algebras  $\mathcal{A}$  such that for every  $s \in S$ ,  $A_s$  is equipped with a preorder  $\geq$ . Hence,  $\mathcal{A} \models \varphi$  if and only if for every variable interpretation  $\nu : X \rightarrow A$ , if each  $\nu(t_i)^{\mathcal{A}} \geq \nu(t'_i)^{\mathcal{A}}$  then  $\nu(t)^{\mathcal{A}} \geq \nu(t')^{\mathcal{A}}$  where  $\_{}^{\mathcal{A}} : T_F(A) \rightarrow A$  is the mapping inductively defined by:  $f(t_1, \dots, t_n)^{\mathcal{A}} = f^{\mathcal{A}}(t_1^{\mathcal{A}}, \dots, t_n^{\mathcal{A}})$ .

**Modal First Order Logic with global satisfaction (MFOL)**<sup>5</sup> The category of signatures is the category of **FOOL** signatures.

Given a **FOOL** signature  $\Sigma = (S, F, P)$ ,  $\Sigma$ -axioms are of the form

<sup>3</sup> $S^+$  is the set of all non-empty sequences of elements in  $S$  and  $S^* = S^+ \cup \{\epsilon\}$  where  $\epsilon$  denotes the empty sequence.

<sup>4</sup> $T_F(X)_s$  is the term algebra of sort  $s$  built over  $F$  with sorted variables in a given set  $X$ .

<sup>5</sup>aka. quantified modal logic **K**.

$p(t_1, \dots, t_n)$  and the set of  $\Sigma$ -formulae is the least set of formulae built over the set of  $\Sigma$ -axioms by finitely applying Boolean connectives in  $\{\neg, \vee\}$  and the quantifier  $\forall$  and the modality  $\Box$ .

Given a signature  $\Sigma = (S, F, P)$ , a  $\Sigma$ -model  $(W, R)$ , called Kripke frame, consists of a family  $W = (W^i)_{i \in I}$  of  $\Sigma$ -models in **FOL** (the possible worlds) such that <sup>6</sup>  $W_s^i = W_s^j$  for every  $i, j \in I$  and  $s \in S$ , and an “accessibility” relation  $R \subseteq I \times I$ . Given a signature morphism  $\sigma : (S, F, P) \rightarrow (S', F', P')$  and a  $(S', F', P')$ -model  $(W', R')$ ,  $\text{Mod}(\sigma)((W', R'))$  is the  $(S, F, P)$ -model  $(W, R)$  defined for every  $i \in I$  by  $W^i = \text{Mod}(\sigma)(W'^i)$  and by  $R = R'$ . A  $\Sigma$ -sentence  $\varphi$  is said to be satisfied by a  $\Sigma$ -model  $(W, R)$ , noted  $(W, R) \models_\Sigma \varphi$ , if for every  $i \in I$  we have  $(W, R) \models_\Sigma^i \varphi$ , where  $\models_\Sigma^i$  is inductively defined on the structure of  $\varphi$  as follows:

- atoms, Boolean connectives and quantifier are handled as in **FOL**,
- $(W, R) \models_\Sigma^i \Box \varphi$  when  $(W, R) \models_\Sigma^j \varphi$  for every  $j \in I$  such that  $i R j$ .

Modal propositional logic (**MPL**) is the sub-institution of **MFOL** whose signatures are restricted to empty sets of sorts and function names and only 0-ary predicate names.

**Modal First Order Logic with local satisfaction (LMFOL)** Signatures and sentences are **MFOL** signatures and **MFOL** sentences. Given a signature  $\Sigma = (S, F, P)$ , a  $\Sigma$ -model is a pointed Kripke frame  $(W = (W^i)_{i \in I}, R, W^j)$  where  $j \in I$ . The satisfaction of a  $\Sigma$ -sentence  $\varphi$  by a  $\Sigma$ -model  $(W, R, W^j)$ , noted  $(W, R, W^j) \models_\Sigma \varphi$ , is defined by:  $(W, R, W^j) \models_\Sigma \varphi \Leftrightarrow (W, R) \models_\Sigma^j \varphi$ .

**LMFOL with infinite disjunction and conjunction (LIMFOL)** This institution extends **LMFOL** to sentences of the form  $\bigwedge \Phi$  and  $\bigvee \Phi$  where  $\Phi$  is a set (possibly infinite) of  $\Sigma$ -sentences. Given a pointed Kripke frame  $(W, R, W^j)$ ,

- $(W, R, W^j) \models_\Sigma \bigwedge \Phi \Leftrightarrow \forall \varphi \in \Phi, (W, R, W^j) \models_\Sigma \varphi$
- $(W, R, W^j) \models_\Sigma \bigvee \Phi \Leftrightarrow \exists \varphi \in \Phi, (W, R, W^j) \models_\Sigma \varphi$

## 2.2 Theories in institutions

Let us now consider a fixed but arbitrary institution  $\mathcal{I} = (\text{Sig}, \text{Sen}, \text{Mod}, \models)$ .

**Notation 2.3** Let  $\Sigma \in |\text{Sig}|$  be a signature and  $T$  be a set of  $\Sigma$ -sentences.

- $\text{Mod}(T)$  is the full sub-category of  $\text{Mod}(\Sigma)$  whose objects are models of  $T$ ,
- $T^\bullet = \{\varphi \in \text{Sen}(\Sigma) / \forall \mathcal{M} \in |\text{Mod}(T)|, \mathcal{M} \models_\Sigma \varphi\}$  is the set of so-called semantic consequences of  $T$ .

---

<sup>6</sup>In the literature, Kripke frames satisfying such a property are said *with constant domains*.

**Definition 2.4 (Theory)** A set  $T$  of  $\Sigma$ -sentences is said to be a theory if and only if  $T = T^\bullet$ .

**Definition 2.5 (The category of theories)** A theory morphism from a  $\Sigma$ -theory  $T$  to a  $\Sigma'$ -theory  $T'$  is any signature morphism  $\sigma : \Sigma \rightarrow \Sigma'$  such that  $Sen(\sigma)(T) \subseteq T'$ .

Let us note  $Th_{\mathcal{I}}$  the category whose objects and morphisms are theories and theory morphisms respectively.

The following proposition is a direct consequence of the satisfaction condition.

**Proposition 2.6** Given a  $\Sigma$ -theory  $T$  and a  $\Sigma'$ -theory  $T'$  such that there is a theory morphism  $\sigma : T \rightarrow T'$ , the functor  $Mod(\sigma) : Mod(\Sigma') \rightarrow Mod(\Sigma)$  can be restricted to  $Mod(\sigma) : Mod(T') \rightarrow Mod(T)$ .

**Corollary 2.7** The 4-uple  $\mathcal{I}_T = (Sig_T, Sen_T, Mod_T, \models_T)$  where:

- $Sig_T = Th_{\mathcal{I}}$ ,
- for every theory  $T$  over a signature  $\Sigma$ ,  $Sen_T(T) = Sen(\Sigma)$  and  $Mod_T(T) = Mod(T)$ , and
- $\models_T = \models$ .

is an institution.

### 2.3 Institution morphisms and comorphisms

Many different kind of morphism can be defined denoting different kind of relationship between two institutions. The original one introduced in [15] defines a forgetful operation from a “richer” institution  $\mathcal{I}$  to a “poorer” one  $\mathcal{I}'$ . Intuitively, it shows how  $\mathcal{I}$  is built over  $\mathcal{I}'$ .

**Definition 2.8 (Institution morphism)** Let  $\mathcal{I} = (Sig, Sen, Mod, \models)$  and  $\mathcal{I}' = (Sig', Sen', Mod', \models')$  be two institutions. An institution-morphism  $\mu = (\Phi, \alpha, \beta) : \mathcal{I} \rightarrow \mathcal{I}'$  consists of

- a functor  $\Phi : Sig \rightarrow Sig'$ ,
- a natural transformation  $\alpha : Sen' \circ \Phi \Rightarrow Sen$ , i.e. for every  $\Sigma \in |Sig|$  a function  $\alpha_\Sigma : Sen'(\Phi(\Sigma)) \rightarrow Sen(\Sigma)$  such that for every signature morphism  $\sigma : \Sigma_1 \rightarrow \Sigma_2$  in  $Sig$  the following diagram commutes,

$$\begin{array}{ccc}
 Sen'(\Phi(\Sigma_1)) & \xrightarrow{\alpha_{\Sigma_1}} & Sen(\Sigma_1) \\
 Sen'(\Phi(\sigma)) \downarrow & & \downarrow Sen(\sigma) \\
 Sen'(\Phi(\Sigma_2)) & \xrightarrow{\alpha_{\Sigma_2}} & Sen(\Sigma_2)
 \end{array}$$



- a natural transformation  $\beta : Mod \Rightarrow Mod' \circ \Phi^{op}$ , i.e. for every  $\Sigma \in |Sig|$  a functor  $\beta_\Sigma : Mod(\Sigma) \rightarrow Mod'(\Phi(\Sigma))$  such that for every signature morphism  $\sigma : \Sigma_1 \rightarrow \Sigma_2$  in  $Sig$  the following diagram commutes,

$$\begin{array}{ccc}
Mod(\Sigma_2) & \xrightarrow{\beta_{\Sigma_2}} & Mod'(\Phi(\Sigma_2)) \\
Mod(\sigma) \downarrow & & \downarrow Mod'(\Phi^{op}(\sigma)) \\
Mod(\Sigma_1) & \xrightarrow{\beta_{\Sigma_1}} & Mod'(\Phi(\Sigma_1))
\end{array}$$

such that the following satisfaction property holds:  
 $\forall \Sigma \in |Sig|, \forall \mathcal{M} \in |Mod(\Sigma)|, \forall \varphi' \in Sen'(\Phi(\Sigma))$

$$\mathcal{M} \models_\Sigma \alpha_\Sigma(\varphi') \iff \beta_\Sigma(\mathcal{M}) \models_{\Phi(\Sigma)} \varphi'$$

**Example 2.9** The institution morphism  $\mu = (\Phi, \alpha, \beta)$  from **FOL** to **EQ**L maps any **FOL** signature  $(S, F, P)$  to the corresponding algebraic one  $(S, F)$ , regards any set of equations as a set of first-order sentences over  $(S, F, \emptyset)$ , and regards any  $(S, F, P)$ -model as a  $(S, F)$ -algebra by forgetting the interpretations of predicate names in  $P$ . It is easy to show that the satisfaction property holds. The institution morphism from **HCL** to **EQ**L is defined as the previous one except that it regards any equation as a conditional equational formula without premises. Finally, the institution morphism from **FOL** to **HCL** is obvious. Indeed, it maps any **FOL** signature  $(S, F, P)$  and any  $(S, F, P)$ -model to themselves, and regards Horn sentences over  $(S, F, P)$  as first-order sentences over the same signature.<sup>7</sup>

No institution morphism can denote a forgetful operation from **FOL** to either **LMFOL** nor **MFOL**. However, we can show how both **LMFOL** and **MFOL** are “embedded” in **FOL**. Indeed, each **LMFOL** signature  $(S, F, P)$  can be transformed into the **FOL** signature  $(\overline{S}, \overline{F}, \overline{P})$  defined by:

- $\overline{S} = S \cup \{ind\}$
- $\overline{F} = \{f : ind \times s_1 \times \dots \times s_n \rightarrow s / f : s_1 \times \dots \times s_n \rightarrow s \in F\} \cup \{i : \rightarrow ind\}$
- $\overline{P} = \{r : ind \times s_1 \times \dots \times s_n / r : s_1 \times \dots \times s_n \in P\} \cup \{R : ind \times ind\}$

Let  $X$  be a set of variables over  $(\overline{S}, \overline{F}, \overline{P})$ , and let  $x \in X_{ind} \cup \{i\}$ . We can define  $FO_x : Sen_{\mathbf{LMFOL}}(\Sigma) \times T_{\overline{F}}(X) \rightarrow Sen_{\mathbf{FOL}}(\overline{\Sigma}) \times T_{\overline{F}}(X)$  inductively on terms and formulae structure as follows:

- $f(t_1, \dots, t_n) \mapsto f(x, t_1, \dots, t_n)$
- $r(t_1, \dots, t_n) \mapsto r(x, t_1, \dots, t_n)$

---

<sup>7</sup>A Horn sentence can be seen as a finite disjunction of first-order formulae all of them being negations of atoms except for the last one which is an atom.

- $\varphi \vee \psi \mapsto FO_x(\varphi) \vee FO_x(\psi)$
- $\neg\varphi \mapsto \neg FO_x(\varphi)$
- $\forall y.\varphi \mapsto \forall y.FO_x(\varphi)$
- $\Box\varphi \mapsto \forall y, x R y \Rightarrow FO_y(\varphi)$

Hence, given a **LMFOL** signature  $(S, F, P)$ , this defines a mapping  $\alpha_{(S, F, P)} : Sen_{\mathbf{LMFOL}}((S, F, P)) \rightarrow Sen_{\mathbf{FOL}}((\overline{S}, \overline{F}, \overline{P}))$  as:  $\alpha_{(S, F, P)}(\varphi) = FO_i(\varphi)$ .

Given a  $(\overline{S}, \overline{F}, \overline{P})$ -model  $\mathcal{M}$ , define the pointed Kripke frame  $\mathcal{W}_{\mathcal{M}} = (W, R, W^i)$  over  $(S, F, P)$  as follows:

- for every  $j \in M_{ind}$ , define the  $(S, F, P)$ -model  $W^j$  as follows:
  - for every  $s \in S$ ,  $W^j_s = M_s$
  - for every  $f : ind \times s_1 \times \dots \times s_n \rightarrow s \in \overline{F} \setminus \{i\}$  and for every  $(a_1, \dots, a_n) \in M_{s_1} \times \dots \times M_{s_n}$ ,  $f^{W^j}(a_1, \dots, a_n) = f^{\mathcal{M}}(j, a_1, \dots, a_n)$ , and
  - for every  $r : ind \times s_1 \times \dots \times s_n \in \overline{P} \setminus \{R\}$  and for every  $(a_1, \dots, a_n) \in M_{s_1} \times \dots \times M_{s_n}$ ,  $(a_1, \dots, a_n) \in r^{W^j} \Leftrightarrow (j, a_1, \dots, a_n) \in r^{\mathcal{M}}$
- $R = R^{\mathcal{M}}$

Hence, given a **LMFOL** signature  $(S, F, P)$ , this defines a mapping  $\beta_{(S, F, P)} : Mod_{\mathbf{FOL}}((\overline{S}, \overline{F}, \overline{P})) \rightarrow Mod_{\mathbf{LMFOL}}((S, F, P))$  which maps any **FOL** model  $\mathcal{M}$  over  $(\overline{S}, \overline{F}, \overline{P})$  to the corresponding **LMFOL** model  $\mathcal{W}_{\mathcal{M}}$  over  $(S, F, P)$ . Notice that the satisfaction of sentences is invariant with respect to this embedding.

Such an embedding relationship between institutions is very well captured by the concept of *institution comorphism*.

**Definition 2.10 (Institution comorphism)** Let  $\mathcal{I} = (Sig, Sen, Mod, \models)$  and  $\mathcal{I}' = (Sig', Sen', Mod', \models')$  be two institutions. An institution-comorphism  $\mu = (\Phi, \alpha, \beta) : \mathcal{I}' \rightarrow \mathcal{I}$  consists of

- a functor  $\Phi : Sig' \rightarrow Sig$ ,
- a natural transformation  $\alpha : Sen' \Rightarrow Sen \circ \Phi$ , and
- a natural transformation  $\beta : Mod \circ \Phi \Rightarrow Mod'$

such that the following satisfaction property holds:

$$\forall \Sigma' \in |Sig'|, \forall \mathcal{M} \in |Mod(\Phi(\Sigma'))|, \forall \varphi' \in Sen'(\Sigma')$$

$$\mathcal{M} \models_{\Phi(\Sigma')} \alpha_{\Sigma'}(\varphi') \iff \beta_{\Sigma'}(\mathcal{M}) \models'_{\Sigma'} \varphi'$$

**Example 2.11** *The institution comorphism from **EQL** to **FOL** maps any algebraic signature  $(S, F)$  to  $(S, F, \emptyset)$  and regards any  $(S, F)$ -equation as a first-order sentence over  $(S, F, \emptyset)$  and any first-order model over  $(S, F, \emptyset)$  as a  $(S, F)$ -algebra. The institution comorphism from **MFOL** to **FOL** is almost the same than the one from **LMFOL** to **FOL** except that  $\overline{F}$  does not contain  $\{i : \rightarrow \text{ind}\}$  anymore. Finally, the institution comorphism  $\mu = (\Phi, \alpha, \beta)$  from **RWL** to **HCL<sub>T</sub>** is defined as follow:*

- $\Phi((S, F)) = T$  where  $T$  is the first-order theory  $(\Sigma', Ax)$  such that:

- $\Sigma' = (S, F, \{\rightarrow_s : s \times s \mid s \in S\})$

- $Ax$  is the following set of  $\Sigma'$ -sentences:

- \* for every  $f : s_1 \times \dots \times s_n \rightarrow s \in F$ ,

$$\bigwedge_{i \leq n} (x_i \rightarrow_{s_i} x'_i) \Rightarrow f(x_1, \dots, x_n) \rightarrow_s f(x'_1, \dots, x'_n)$$

- \*  $x \rightarrow_s x$

- \*  $x \rightarrow_s y \wedge y \rightarrow_s z \Rightarrow x \rightarrow_s z$

- $\alpha_\Sigma(\bigwedge_{j \leq n} t_j \rightarrow t'_j \Rightarrow t_{n+1} \rightarrow t'_{n+1}) = \bigwedge_{j \leq m} t_j \rightarrow_{s_j} t'_j \Rightarrow t_{n+1} \rightarrow_{s_{n+1}} t'_{n+1}$

- $\beta_\Sigma(\mathcal{M}) = \mathcal{A}$  where  $\mathcal{A}$  is the  $\Sigma$ -algebra such that for every  $s \in S$ ,  $A_s = M_s$  and is equipped with preorder  $\rightarrow_s^{\mathcal{M}}$ , and for every  $f \in F$ ,  $f^{\mathcal{A}} = f^{\mathcal{M}}$ .

### 3 Craig interpolation property

#### 3.1 The institution-independent formulation

We follow the institution-independent formulation of the Craig interpolation property for sentences given in [8]. This formulation generalises the conventional intersection-union (of signatures) framework to any square of signature morphisms. This formulation also generalises previous institution-independent ones more restrictive (squares have to be pushout [5, 11] or abstract inclusions [10]).

**Definition 3.1 (Craig interpolation square)** *A commutative square of signature morphisms*

$$\begin{array}{ccc} \Sigma & \xrightarrow{\sigma_1} & \Sigma_1 \\ \sigma_2 \downarrow & & \downarrow \sigma'_1 \\ \Sigma_2 & \xrightarrow[\sigma'_2]{} & \Sigma' \end{array}$$

*is a Craig interpolation square if and only if for every set  $E_1$  of  $\Sigma_1$ -sentences and set  $E_2$  of  $\Sigma_2$ -sentences such that  $\text{Sen}(\sigma'_1)(E_1) \models_{\Sigma'} \text{Sen}(\sigma'_2)(E_2)$ , there exists a*

set  $E$  of  $\Sigma$ -sentences such that  $E_1 \models_{\Sigma_1} \text{Sen}(\sigma_1)(E)$  and  $\text{Sen}(\sigma_2)(E) \models_{\Sigma_2} E_2$ . The set  $E$  is called an interpolant of  $E_1$  and  $E_2$ .

An institution  $\mathcal{I}$  has the Craig interpolation property (CIP) if and only if every commutative square of signature morphisms is a Craig interpolation square.

**Example 3.2** While **PL** and the unsorted sub-institutions of **FOL**, **LMFOL** and **LIMFOL** have CIP for every commutative square [2, 11], (many-sorted) **FOL**, **LMFOL** and **LIMFOL** only have CIP for commutative squares with one component injective on sorts [2, 16]. **HCL** and **EQ** only have CIP for commutative squares with  $\sigma_2$  injective [8]. Finally, it is well-known that CIP fails [12] for the quantified modal logic **K** with constant domains.<sup>8</sup> Hence, **MFOL** doesn't have CIP.

### 3.2 Preservation of CIP

It may sometimes be difficult, or simply tedious, to prove CIP for a particular institution. If there exists another institution for which CIP has already been shown or is easier to prove and which is linked to the former one by a particular notion of morphism, it is tempting to prove CIP through this morphism. The following theorem establishes such a result for institution morphisms.

**Theorem 3.3** Let  $\mu = (\Phi, \alpha, \beta) : \mathcal{I} \rightarrow \mathcal{I}'$  be an institution morphism such that there is a functor  $\overline{\Phi} : \text{Sig}' \rightarrow \text{Sig}$  satisfying  $\Phi \circ \overline{\Phi} = \text{Id}_{\text{Sig}'}$  and for all  $\Sigma \in |\text{Sig}|$ ,  $\beta_\Sigma$  and  $\alpha_\Sigma$  are surjective. If  $\mathcal{I}$  has CIP then so does  $\mathcal{I}'$  as well.

**Proof** Let

$$\begin{array}{ccc} \Sigma' & \xrightarrow{\sigma_1} & \Sigma'_1 \\ \sigma_2 \downarrow & & \downarrow \sigma'_1 \\ \Sigma'_2 & \xrightarrow[\sigma'_2]{} & \Sigma'' \end{array}$$

be a commutative square of signature morphisms in  $\text{Sig}'$ . Let  $E'_1 \subseteq \text{Sen}'(\Sigma'_1)$  and  $E'_2 \subseteq \text{Sen}'(\Sigma'_2)$  such that  $\text{Sen}(\sigma'_1)(E'_1) \models'_{\Sigma''} \text{Sen}(\sigma'_2)(E'_2)$ . Let  $\mathcal{M} \in \text{Mod}(\alpha_{\overline{\Phi}(\Sigma'')})$ . As  $\mu$  is an institution morphism, by the satisfaction property we have:

$$\begin{aligned} \mathcal{M} \models_{\overline{\Phi}(\Sigma'')} \alpha_{\overline{\Phi}(\Sigma'')}(\text{Sen}'(\sigma'_1)(E'_1)) &\iff \beta_{\overline{\Phi}(\Sigma'')}(\mathcal{M}) \models'_{\Sigma''} \text{Sen}'(\sigma'_1)(E'_1) \\ &\iff \beta_{\overline{\Phi}(\Sigma'')}(\mathcal{M}) \models'_{\Sigma''} \text{Sen}'(\sigma'_2)(E'_2) \\ &\iff \mathcal{M} \models_{\overline{\Phi}(\Sigma'')} \alpha_{\overline{\Phi}(\Sigma'')}(\text{Sen}'(\sigma'_2)(E'_2)) \end{aligned}$$

Hence, we have  $\alpha_{\overline{\Phi}(\Sigma'')}(\text{Sen}'(\sigma'_1)(E'_1)) \models_{\overline{\Phi}(\Sigma'')} \alpha_{\overline{\Phi}(\Sigma'')}(\text{Sen}'(\sigma'_2)(E'_2))$ . As  $\mathcal{I}$  has CIP, there exists  $E \subseteq \text{Sen}(\overline{\Phi}(\Sigma'))$  such that  $\alpha_{\overline{\Phi}(\Sigma'_1)}(E) \models_{\overline{\Phi}(\Sigma'_1)} \text{Sen}(\overline{\Phi}(\sigma_1))(E)$

<sup>8</sup>Actually, CIP fails for a number of quantified modal logics with constant domains such as **T**, **D**, **S4**, and **S5** [12]. It even fails for **S5** with varying domains [12, 13].

and  $\text{Sen}(\overline{\Phi}(\sigma_2))(E) \models_{\overline{\Phi}(\Sigma_2)} \alpha_{\overline{\Phi}(\Sigma_2)}(E'_2)$ . As  $\alpha_\Sigma$  is surjective for every signature  $\Sigma \in |\text{Sig}|$ , the set  $E' = \alpha_{\overline{\Phi}(\Sigma')}^{-1}(E)$  is well defined. Let  $\mathcal{M}' \in \text{Mod}(E'_1)$ . As  $\beta_{\overline{\Phi}(\Sigma'_1)}$  is surjective, there exists  $\mathcal{M} \in \text{Mod}(\overline{\Phi}(\Sigma'_1))$  such that  $\beta_{\overline{\Phi}(\Sigma'_1)}(\mathcal{M}) = \mathcal{M}'$ . Now, by the satisfaction property of institution morphisms, we have:

$$\begin{aligned} \mathcal{M}' \models'_{\Sigma'_1} E'_1 &\iff \mathcal{M} \models_{\overline{\Phi}(\Sigma'_1)} \alpha_{\overline{\Phi}(\Sigma'_1)}(E'_1) \\ &\iff \mathcal{M} \models_{\overline{\Phi}(\Sigma'_1)} \text{Sen}(\overline{\Phi}(\sigma_1))(\alpha_{\overline{\Phi}(\Sigma')} (E')) \\ &\iff \mathcal{M}' \models'_{\Sigma'_1} \text{Sen}'(\sigma_1)(E') \end{aligned}$$

Hence,  $E'_1 \models'_{\Sigma'_1} \text{Sen}'(\sigma_1)(E')$ . Similarly, we prove that  $\text{Sen}'(\sigma_2)(E') \models'_{\Sigma'_2} E'_2$ .  $\square$

By observing the proof of Theorem 3.3, we have the following corollary:

**Corollary 3.4** *With the notations and hypothesis of Theorem 3.3,  $\mathcal{I}'$  has CIP for a commutative diagram  $D$  if  $\mathcal{I}$  has CIP for  $D_\mu$  (the image of  $D$  by  $\overline{\Phi}$ ).*

Dually, for institution comorphisms, a similar result can be stated except that  $\overline{\Phi}$  is of course removed:

**Theorem 3.5** *Let  $\mu = (\Phi, \alpha, \beta) : \mathcal{I}' \rightarrow \mathcal{I}$  be an institution comorphism such that for all  $\Sigma \in |\text{Sig}|$ ,  $\beta_\Sigma$  and  $\alpha_\Sigma$  are surjective. Then, if  $\mathcal{I}$  has CIP then so does  $\mathcal{I}'$  as well.*

The hypothesis that  $\beta_\Sigma$  is surjective is quite natural. Besides, all the examples of institution morphisms and comorphisms given in this paper satisfy such a property. On the contrary, the surjectivity of  $\alpha_\Sigma$  is satisfied by none of the examples of institution morphisms and comorphisms given in this paper except **RWL**  $\rightarrow$  **HCL<sub>T</sub>**. Actually, surjectivity of  $\alpha_\Sigma$  is above all satisfied by institution morphisms and comorphisms where the sentences of the source and the target have the same expressive power (e.g. **FOEQL**  $\rightarrow$  **FOL** embedding the first-order equational logic into the first-order logic). Below, we give a weaker condition, called *conservative* institution morphism and comorphism, defined as follow:

**Notation 3.6** *Let  $\mathcal{I}$  be an institution. Let  $\sigma : \Sigma_1 \rightarrow \Sigma_2$  be a signature morphism and let  $E \subseteq \text{Sen}(\Sigma_2)$ . Let us note  $E_\sigma = \{\varphi \in \text{Sen}(\Sigma_1) / E \models_{\Sigma_2} \text{Sen}(\sigma)(\varphi)\}$ .*

**Definition 3.7 (Conservativeness)** *Let  $\mu = (\Phi, \alpha, \beta) : \mathcal{I}' \rightarrow \mathcal{I}$  be an institution morphism such that there is a functor  $\overline{\Phi} : \text{Sig}' \rightarrow \text{Sig}$  satisfying  $\Phi \circ \overline{\Phi} = \text{Id}_{\text{Sig}'}$ . It is conservative if and only if every signature morphism  $\sigma : \Sigma \rightarrow \Sigma'$  and every set of sentences  $E' \subseteq \text{Sen}'(\Sigma')$  in the institution  $\mathcal{I}'$  satisfy:*

$$\text{Mod}(\alpha_{\overline{\Phi}(\Sigma)}(E'_\sigma)) \subseteq \text{Mod}(\alpha_{\overline{\Phi}(\Sigma')} (E'))_{\overline{\Phi}(\sigma)}$$

An obvious sufficient condition to obtain conservative institution morphisms (resp. comorphisms) is  $\alpha_{\overline{\Phi}(\Sigma')} (E')_{\overline{\Phi}(\sigma)} \subseteq \alpha_{\overline{\Phi}(\Sigma)} (E'_\sigma)$ . This last inclusion is exactly the notion defined in [18] to obtain preservation of CIP along institution transformations and called *restriction adequateness*. The interest of the notion of conservativeness lies in the fact that it is a lesser restrictive property than restriction adequateness. Indeed there are conservative institution morphisms and comorphisms that are not restriction adequate (see Example 3.8 below). The opposite inclusion, *i.e.*  $Mod(\alpha_{\overline{\Phi}(\Sigma')} (E')_{\overline{\Phi}(\sigma)}) \subseteq Mod(\alpha_{\overline{\Phi}(\Sigma)} (E'_\sigma))$ , is obviously satisfied. Indeed, it is obvious to show that  $\alpha_{\overline{\Phi}(\Sigma)} (E'_\sigma) \subseteq \alpha_{\overline{\Phi}(\Sigma')} (E')_{\overline{\Phi}(\sigma)}$ .

**Example 3.8** Consider the institution morphism  $\mathbf{HCL} \rightarrow \mathbf{EQL}$  equipped with the functor  $\overline{\Phi}$  which maps any algebraic signature  $(S, F)$  to  $(S, F, \emptyset)$ . Observe that for this institution morphism, for every algebraic signature  $\Sigma$ ,  $\alpha_\Sigma$  is not surjective. Show that it is conservative. Let us consider a signature morphism  $\sigma : \Sigma \rightarrow \Sigma'$  in  $\mathbf{EQL}$  and a set of  $\Sigma'$ -equations  $E'$ . Horn clauses whose predicates are equations, are classically called conditional positive equations. For such formulae, there exists a complete calculus which extends G. Birkhoff's equational reasoning to conditional equations. Given an algebraic signature  $\Sigma$  and a set of  $\Sigma$ -equations, it is easy to show that the set of theorems  $E^\bullet$  obtained by using the conditional equational reasoning are

- either tautologies of the form  $\bigwedge_{i \leq n} t_i = t'_i \Rightarrow t_j = t'_j$  such that  $j \in \{1, \dots, n\}$ ,
- or conditional equations of the form  $\bigwedge_{i \leq n} t_i = t'_i \Rightarrow t = t'$  for which there exists  $u = v \in E$ , a context  $C$  and a substitution  $\sigma$  such that  $t = C[\sigma(u)]$  and  $t' = C[\sigma(v)]$  or  $t = C[\sigma(v)]$  and  $t' = C[\sigma(u)]$ .<sup>9</sup>

Hence, given an algebra  $\mathcal{M} \in Mod(\alpha_{\overline{\Phi}(\Sigma)} (E'_\sigma))$  and a formula  $\bigwedge_{i \leq n} t_i = t'_i \Rightarrow t = t' \in \alpha_{\overline{\Phi}(\Sigma')} (E')_{\overline{\Phi}(\sigma)}$ , we necessarily have that either  $t = t' \in \{t_1 = t'_1, \dots, t_n = t'_n\}$  or  $t = t' \in \alpha_{\overline{\Phi}(\Sigma)} (E'_\sigma)$ . In both cases, we conclude that  $\mathcal{M} \models_\Sigma \bigwedge_{i \leq n} t_i = t'_i \Rightarrow t = t'$ .

The same condition can be defined for institution comorphism by replacing  $\overline{\Phi}(\_)$  by  $\_$ .

**Theorem 3.9** Let  $\mu = (\Phi, \alpha, \beta) : \mathcal{I} \rightarrow \mathcal{I}'$  be a conservative institution morphism such that for every signature  $\Sigma$  in  $\mathcal{I}$ ,  $\beta_\Sigma$  is surjective. Then, if  $\mathcal{I}$  has CIP then so does  $\mathcal{I}'$  as well.

**Proof** Let

<sup>9</sup>A context  $C$  is a term with a unique occurrence of a constant  $\square$ , and  $C[t]$  denotes the result of replacing in  $C$  the occurrence of  $\square$  by  $t$ .

$$\begin{array}{ccc}
\Sigma' & \xrightarrow{\sigma_1} & \Sigma'_1 \\
\sigma_2 \downarrow & & \downarrow \sigma'_1 \\
\Sigma'_2 & \xrightarrow{\sigma'_2} & \Sigma''
\end{array}$$

be a commutative square of signature morphisms in  $Sig'$ . Let  $E'_1 \subseteq Sen'(\Sigma'_1)$  and  $E'_2 \subseteq Sen'(\Sigma'_2)$  such that  $Sen'(\sigma'_1)(E'_1) \models'_{\Sigma''} Sen'(\sigma'_2)(E'_2)$ . Let us set  $E' = E'_1|_{\sigma_1}$ . By the satisfaction condition of institutions, we have  $E'_1 \models'_{\Sigma'_1} Sen'(\sigma_1)(E')$ . Therefore, let us show that  $Sen'(\sigma_2)(E') \models'_{\Sigma'_2} E'_2$ .

Let  $\mathcal{M}' \in Mod'(Sen'(\sigma_2)(E'))$ . By the satisfaction condition of institution morphisms, for every  $\mathcal{M} \in |Mod(\overline{\Phi}(\Sigma'_2))|$  such that  $\beta_{\overline{\Phi}(\Sigma'_2)}(\mathcal{M}) = \mathcal{M}'$ ,  $\mathcal{M} \models_{\overline{\Phi}(\Sigma'_2)} \alpha_{\overline{\Phi}(\Sigma'_2)}(Sen'(\sigma_2)(E'))$ . By the surjectivity of  $\beta_{\overline{\Phi}(\Sigma'_2)}$ , such a model  $\mathcal{M}$  exists.

By following the same first steps of the proof of Theorem 3.3, we have  $\alpha_{\overline{\Phi}(\Sigma'')}(Sen'(\sigma'_1)(E'_1)) \models_{\overline{\Phi}(\Sigma'')} \alpha_{\overline{\Phi}(\Sigma'')}(Sen'(\sigma'_2)(E'_2))$ .  $\mathcal{I}$  having CIP, there exists  $E \subseteq Sen(\overline{\Phi}(\Sigma'))$  such that both  $\alpha_{\overline{\Phi}(\Sigma'_1)}(E'_1) \models_{\overline{\Phi}(\Sigma'_1)} Sen(\overline{\Phi}(\sigma_1))(E)$  and  $Sen(\overline{\Phi}(\sigma_2))(E) \models_{\overline{\Phi}(\Sigma'_2)} \alpha_{\overline{\Phi}(\Sigma'_2)}(E'_2)$ . Hence,  $E \subseteq \alpha_{\overline{\Phi}(\Sigma')}(E'_1)|_{\overline{\Phi}(\sigma_1)}$ . As  $\mu$  is conservative, we have  $Mod(\alpha_{\overline{\Phi}(\Sigma')}(E')) \subseteq Mod(E)$ , hence  $Sen(\overline{\Phi}(\sigma_2))(\alpha_{\overline{\Phi}(\Sigma')}(E')) \models_{\overline{\Phi}(\Sigma'_2)} \alpha_{\overline{\Phi}(\Sigma'_2)}(E'_2)$ . We conclude that  $\mathcal{M} \models_{\overline{\Phi}(\Sigma'_2)} \alpha_{\overline{\Phi}(\Sigma'_2)}(E'_2)$  and, by satisfaction condition of institution morphisms, that  $\mathcal{M}' \models'_{\Sigma'_2} E'_2$ .  $\square$

As previously, a similar result can be stated for conservative institution comorphisms. Moreover, by observing the proof of Theorem 3.9, we have the following corollary:

**Corollary 3.10** *With the notations and hypothesis of Theorem 3.9,  $\mathcal{I}'$  has CIP for a commutative diagram  $D$  if  $\mathcal{I}$  has CIP for  $D_\mu$  (the image of  $D$  for  $\overline{\Phi}$ ).*

**Example 3.11** *By the result established in Example 3.8, we directly have that **EQL** has CIP for every commutative diagram with  $\sigma_2$  injective. In this case however, the direct proof of CIP in **EQL** give a stronger result, requiring injectivity only on sorts (see Ex. 3.2).*

## 4 Beth definability theorem

The Beth definability theorem establishes an equivalence between the two notions of explicit and implicit definability. Explicit definability is a syntactical notion while implicit definability is a semantical one. Since both notions deal with language inclusion, we first have to define the notion of inclusion between signatures in order to generalise them in the institutional setting. This requires to give a meaning of the set-theoretical notion of inclusion within category theory as institutions only see signatures as objects of a specific category. To reach this purpose, we use the notion of strongly inclusive category given in [10]. Such

a formalization allows us to categorically denote both set-theoretical notions of union and intersection, thus inclusion.

## 4.1 Categorical meaning of set-theoretical operations

We give a categorical meaning of the four usual set-theoretical operations: inclusion, union, intersection and difference. The first three operations have been defined in [10]. The last one has been defined in [1].

**Definition 4.1** *A category  $\mathcal{C}$  is strongly inclusive if and only if  $\mathcal{C}$  is a category with pullbacks and there exists two sub-categories  $I$  and  $E$  of  $\mathcal{C}$  such that:*

1.  $|I| = |E| = |\mathcal{C}|$ ;
2. any  $e \in \text{Hom}_E$  is epic;
3. every morphism  $f$  in  $\mathcal{C}$  can be factored uniquely as  $\iota \circ e$  with  $\iota \in I$  and  $e \in E$ ;
4.  $\text{Hom}_I$  defines a partial order on  $\mathcal{C}$  and the poset  $(|\mathcal{C}|, \text{Hom}_I)$  is a lattice where the sup of  $A$  and  $B$  is the sum of  $A$  and  $B$ , denoted  $A + B$ , and the inf of  $A$  and  $B$  is the unique inclusion pullback in  $\mathcal{C}$  of the sum, called intersection of  $A$  and  $B$  and denoted  $A \cap B$  (both existence and uniqueness of  $A \cap B$  are given in [10]). Moreover, we impose on  $A + B$  to be the pushout of  $A \cap B$ .

The morphisms in  $I$  are called inclusions and the pair  $(I, E)$  is called the inclusion system of  $\mathcal{C}$ . In the following we will use the notation  $A \hookrightarrow B$  to denote the morphism  $\iota : A \rightarrow B$  in  $\text{Hom}_I(A, B)$ .

**Remark 4.2** *Sum and intersection are unique and the full sub-category whose morphisms are inclusions is finitely cocomplete (see [10] for the complete proofs of these results).*

**Example 4.3** *As an example, the category  $\text{Set}$  provided with the inclusion system  $(I, E)$  where  $I$  contains all inclusions and  $E$  contains all surjections is*

*trivially strongly inclusive. Indeed, given a pushout*

$$\begin{array}{ccc}
 s_0 & \xrightarrow{g_1} & s_1 \\
 g_2 \downarrow & & \downarrow \\
 & & s
 \end{array}$$

*in  $\text{Set}$  with  $S_0 = S_1 \cap S_2$  and all arrows inclusions,  $S$  denotes the set-theoretical union of  $S_1$  and  $S_2$  (i.e.  $S = S_1 \cup S_2$ ). This pushout is also a pullback, so  $S_1 \cup S_2$  is the pushout of  $S_1 \cap S_2$ .*

**Fact 4.4** *Let  $\mathcal{C}$  be a strongly inclusive category the inclusion system of which is  $(I, E)$ . If  $\mathcal{C}$  has an initial object  $\emptyset$ , then  $\emptyset$  is also initial in  $I$ .*

As argued in [10], even though set inclusions are the simplest natural example of inclusion system, inclusion systems in general may have properties that are



quite different from those of sets. For example, any inclusion  $A \hookrightarrow B$  in set theory is "split" in the sense that  $B$  can be written as a disjoint union  $A \amalg C$  where  $C$  is the "difference" between  $B$  and  $A$ . This property does not hold for every inclusion systems. For example, given a many sorted signature inclusion  $(\{s\}, \emptyset) \hookrightarrow (\{s\}, \{f : s \rightarrow s\})$ , the difference  $(\emptyset, \{f : s \rightarrow s\})$  is not a signature. The expected result is  $(\emptyset, \emptyset)$ . Nevertheless, this notion can be expressed in a categorical manner if we do not ground this axiomatization on the notion of disjoint union. Indeed, we can characterize the difference between  $B$  and  $A$ , denoted  $B \setminus A$ , as the greatest set  $X$  included in  $B$  such that  $X \cap A = \emptyset$ . Therefore, we assume in the following that every strongly inclusive category  $\mathcal{C}$  satisfies the following supplementary condition:

**Definition 4.5 (Categorical difference)** *Let  $\mathcal{C}$  be a strongly inclusive category with an initial object  $\emptyset$ .  $\mathcal{C}$  satisfies the property of difference if and only if for every object  $C$  the endofunctor  $-\setminus C : I \rightarrow I$  that maps the object  $C'$  to the terminal object  $C' \setminus C$  of the full sub-category of  $I$ , the object class of which is  $\{X/X \cap C = \emptyset \wedge X \hookrightarrow C'\}$ , exists. For any two objects  $C$  and  $C'$  we call the object  $C' \setminus C$  the difference between  $C'$  and  $C$ .*

**Notation 4.6** *In a category  $\mathcal{C}$  we note  $C \approx C'$  to mean that  $C$  and  $C'$  are isomorphic objects, that is there is two morphisms  $\alpha : C \rightarrow C'$  and  $\alpha^{-1} : C' \rightarrow C$  such that  $\alpha \circ \alpha^{-1} = Id_{C'}$  and  $\alpha^{-1} \circ \alpha = Id_C$ .*

**Fact 4.7** *In any strong inclusive category  $\mathcal{C}$  that satisfies the property of difference, for every  $C \in |\mathcal{C}|$ ,  $-\setminus C$  preserves isomorphism.*

**Proof** Obvious since every functor preserves isomorphism.  $\square$

In the sequel, we assume that the institution  $\mathcal{I} = (Sig, Sen, Mod, \models)$  satisfies both following supplementary conditions:

1.  $Sig$  is a strongly inclusive category whose associated inclusion system is  $(I, E)$ . It has an initial object, denoted  $\Sigma_\emptyset$ , which is also initial in  $I$ . Finally,  $Sig$  satisfies the property of difference.
2.  $\forall (\Sigma \hookrightarrow \Sigma'), Sen(\Sigma) \subseteq Sen(\Sigma')$

Moreover, we assume that for every institution morphism (resp. comorphism)  $\mu = (\Phi, \alpha, \beta)$ ,  $\Phi$  preserves the inclusion system.

All the examples of institutions given in this paper satisfy the above properties.

## 4.2 The institution-independent formulation

**Definition 4.8 (Elementary equivalence)** *Let  $\mathcal{I} = (Sig, Sen, Mod, \models)$  be an institution. Let  $\Sigma$  be a signature. Two  $\Sigma$ -models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  are elementary equivalent, noted  $\mathcal{M}_1 \equiv_\Sigma \mathcal{M}_2$  if and only if the following condition holds:*

$$\forall \varphi \in Sen(\Sigma), \mathcal{M}_1 \models_\Sigma \varphi \iff \mathcal{M}_2 \models_\Sigma \varphi$$

**Definition 4.9 (Implicitly definable)** Let  $\Sigma \hookrightarrow \Sigma'$  be a signature inclusion and let  $T$  be a  $\Sigma'$ -theory.  $\Sigma' \setminus \Sigma$  is implicitly definable by  $T$  if and only if for all  $\mathcal{M}'$  and  $\mathcal{M}''$  in  $Mod(T)$ , if  $Mod(\Sigma \hookrightarrow \Sigma')(\mathcal{M}') = Mod(\Sigma \hookrightarrow \Sigma')(\mathcal{M}'')$ , then  $\mathcal{M}' \equiv_{\Sigma'} \mathcal{M}''$ .

Our notion of implicit definability is more liberal than the one from conventional model theory. Only the elementary equivalence is required between both  $T$ -models  $\mathcal{M}'$  and  $\mathcal{M}''$  whereas in the conventional model theory both  $T$ -models  $\mathcal{M}'$  and  $\mathcal{M}''$  need to be equal. In the institutional language, this means that the reduct functor  $Mod(\Sigma \hookrightarrow \Sigma') : Mod(T) \rightarrow Mod(\Sigma, \emptyset)$  is injective.

**Definition 4.10 (Explicitly definable)** Let  $\Sigma \hookrightarrow \Sigma'$  and  $T$  be a  $\Sigma'$ -theory.  $\Sigma' \setminus \Sigma$  is explicitly definable relative to  $T$  if for every  $\varphi \in Sen(\Sigma' \setminus \Sigma)$  there exists  $E \subseteq Sen(\Sigma)$  such that both following conditions are satisfied:

1.  $T \cup \{\varphi\} \models_{\Sigma'} E$
2.  $T \cup E \models_{\Sigma'} \varphi$

**Definition 4.11 (Beth property)** An institution  $\mathcal{I}$  has the Beth property (BP) if and only if for all  $\Sigma \hookrightarrow \Sigma'$  and all  $\Sigma'$ -theory  $T$ , if  $\Sigma' \setminus \Sigma$  is implicitly definable by  $T$ , then  $\Sigma' \setminus \Sigma$  is explicitly definable relative to  $T$ . We then say that  $\Sigma' \setminus \Sigma$  has the Beth property for  $T$ .

In conventional model theory, the Beth theorem establishes a correspondence between implicitly and explicitly definable properties. In such a theory, showing that the explicitly definable property implies the implicitly one is quite direct thanks to formula's inductive structure. Indeed, by compactness, we can show that for any  $\varphi \in Sen(\Sigma' \setminus \Sigma)$  ( $\Sigma'$  and  $\Sigma$  are first-order signatures), the set  $E$  in Definition 4.10 is restricted to a formula  $\psi_\varphi \in Sen(\Sigma)$ . Therefore, we can build for any formula  $\Phi \in Sen(\Sigma')$  a formula  $\Pi \in Sen(\Sigma)$  such that  $Mod(T \cup \{\Phi\}) = Mod(T \cup \{\Pi\})$ .  $\Pi$  is obtained by replacing any occurrence of any atomic formula  $p'(t_1, \dots, t_n) \in Sen(\Sigma' \setminus \Sigma)$  by  $\psi_{p'(t_1, \dots, t_n)} \in Sen(\Sigma)$ . Hence, for all  $\mathcal{M}', \mathcal{M}'' \in Mod(T)$  if  $Mod(\Sigma \hookrightarrow \Sigma')(\mathcal{M}') = Mod(\Sigma \hookrightarrow \Sigma')(\mathcal{M}'')$ , then  $\mathcal{M}' \equiv_{\Sigma'} \mathcal{M}''$ . Indeed, for every  $\Phi \in Sen(\Sigma')$  we have:

$$\begin{aligned} \mathcal{M}' \models_{\Sigma'} \Phi &\iff \mathcal{M}' \models_{\Sigma'} \Pi \\ &\iff Mod(\Sigma \hookrightarrow \Sigma')(\mathcal{M}') \models_{\Sigma} \Pi \\ &\iff \mathcal{M}'' \models_{\Sigma'} \Pi \\ &\iff \mathcal{M}'' \models_{\Sigma'} \Phi \end{aligned}$$

In the institution framework, in order to ensure that explicit definability implies implicit one, we have to extend the explicit definability notion by replacing in Definition 4.10 the sentence “for every  $\varphi \in Sen(\Sigma' \setminus \Sigma)$ ” by “for every  $\varphi \in Sen(\Sigma') \setminus Sen(\Sigma)$ ”. From this new notion of explicit definability, we have the expected result, that is:

**Proposition 4.12 (Explicit implies implicit)** If  $\Sigma' \setminus \Sigma$  is explicitly definable relative to  $T$  then it is implicitly for  $T$ .

**Proof** Let  $\mathcal{M} \in |Mod(\Sigma)|$  and  $\mathcal{M}_i \in Mod(T)$  such that  $Mod(\Sigma \leftrightarrow \Sigma')(\mathcal{M}_i) = \mathcal{M}$  ( $i = 1, 2$ ). Let  $\varphi \in Sen(\Sigma')$  such that  $\mathcal{M}_i \models_{\Sigma'} \varphi$ . Let us show that  $\mathcal{M}_j \models_{\Sigma'} \varphi$  ( $j \neq i, i, j \in \{1, 2\}$ ). By the satisfaction condition, this is obvious if  $\varphi \in Sen(\Sigma)$ . Therefore, suppose that  $\varphi \in Sen(\Sigma') \setminus Sen(\Sigma)$ . As  $\Sigma' \setminus \Sigma$  is explicitly definable relative to  $T$ , there exists  $E$  such that  $T \cup E \models_{\Sigma'} \varphi$  and  $T \cup \{\varphi\} \models_{\Sigma'} E$ . Suppose that  $\mathcal{M}_j \not\models_{\Sigma'} \varphi$ . This means that there is some  $\psi \in E$  such that  $\mathcal{M}_j \not\models_{\Sigma'} \psi$ . By the satisfaction condition, we then have  $\mathcal{M} \not\models_{\Sigma'} \psi$ . But, from  $T \cup \{\varphi\} \models E$  and the satisfaction condition, we also have that  $\mathcal{M} \models_{\Sigma'} \psi$  which is impossible.  $\square$

In the institution framework, this new notion of explicit definability is not implied by the one defined in Definition 4.10. The reason is that formulae are just elements of a set. However, the proof given just above for the first-order logic may let us suppose that such a result will hold in abstract frameworks of logics with inductively defined sets of formulae such as Goguen and Burstall's parchments [14] or Fiblog group's abstract logics (e.g. [19]). Moreover, in the conventional model theory, the Beth property is defined for atomic formulae. This is enough to generalise this property for more general signature inclusions. Indeed, for the first-order logic, the following result holds:

**Proposition 4.13** *In first-order logic, if  $\Sigma' \setminus \Sigma$  and  $\Sigma'' \setminus \Sigma$  are explicitly definable for  $T'$  and  $T''$  respectively, then so does  $\Sigma' + \Sigma'' \setminus \Sigma$  for  $(T' \cup T'')^\bullet$ .*

**Proof** Suppose  $\varphi \in Sen(\Sigma' + \Sigma'') \setminus Sen(\Sigma)$ . By induction on the structure of  $\varphi$ , let us show that there exists  $E \subseteq Sen(\Sigma)$  such that  $T \cup E \models_{\Sigma' + \Sigma''} \varphi$  and  $T \cup \{\varphi\} \models_{\Sigma' + \Sigma''} E$  where  $T = (T' \cup T'')^\bullet$ .

- *basic case:* directly results from the explicit definability of  $\Sigma' \setminus \Sigma$  and  $\Sigma'' \setminus \Sigma$  relative to  $T'$  and  $T''$  respectively.
- *general case:* As previously, suppose that the only connectors and quantifier are  $\neg$ ,  $\wedge$  and  $\forall$ , respectively.
  - $\varphi$  is of the form  $\neg\psi$ . By induction hypothesis, there exists  $E'$  such that  $T \cup E' \models \psi$  and  $T \cup \{\psi\} \models E'$ . By compactness and the deduction lemma, there exists a finite subset  $\{\xi_1, \dots, \xi_n\} \subset E'$  such that  $T \models \psi \Leftrightarrow \bigwedge_{1 \leq i \leq n} \xi_i$ . It is sufficient to set  $E = (E' \setminus \{\xi_1, \dots, \xi_n\}) \cup \{\neg \bigwedge_{1 \leq i \leq n} \xi_i\}$ .
  - $\varphi$  is of the form  $\psi_1 \wedge \psi_2$ . By induction hypothesis, there exists  $E_i$  such that  $T \cup E_i \models \psi_i$  and  $T \cup \{\psi_i\} \models E_i, i = 1, 2$ . It is sufficient to set  $E = E_1 \cup E_2$ .
  - $\varphi$  is of the form  $\forall x.\psi$ . By induction hypothesis, there exists  $E'$  such that  $T \cup E' \models \psi$  and  $T \cup \{\psi\} \models E'$ . It is sufficient to set  $E = E'$ .

□

The similar result for the implicit case does not hold when both first-order structures  $\mathcal{M}_i$  and  $\mathcal{M}_j$  are required to be elementary equivalent (this is due to negation). However in conventional model theory, the implicit definability notion only requires that  $\mathcal{M}_i$  and  $\mathcal{M}_j$  are equal. In this case, the previous result can be extended to the case of implicit definability.

### 4.3 The institution-independent proof

**Lemma 4.14** *Let  $\Sigma' \setminus \Sigma$  be a signature that is implicitly definable by a  $\Sigma'$ -theory  $T$ . Let  $\Sigma \hookrightarrow \Sigma''$  be an inclusion morphism such that  $\Sigma' \approx \Sigma''$ . Let us note  $\alpha : \Sigma' \rightarrow \Sigma$  the isomorphism associated with  $\Sigma' \approx \Sigma''$ . Then, for every  $E \subseteq \text{Sen}(\Sigma' \setminus \Sigma)$ ,  $T \cup E \models_{\Sigma' + \Sigma''} \text{Sen}(\alpha)(T \cup E)$ .*

**Proof** Let  $\mathcal{M} \models_{\Sigma' + \Sigma''} T \cup E$ . By Proposition 2.6,  $\text{Mod}(\Sigma' \hookrightarrow \Sigma' + \Sigma'')(\mathcal{M}) \models_{\Sigma'} \text{Sen}(\alpha^{-1} \circ \alpha)(T \cup E)$ , and then  $\text{Mod}(\alpha^{-1})(\text{Mod}(\Sigma' \hookrightarrow \Sigma' + \Sigma'')(\mathcal{M})) \models_{\Sigma''} \text{Sen}(\alpha)(T \cup E)$ . Obviously, we have  $\Sigma'' \hookrightarrow \Sigma' + \Sigma'' \circ \alpha \circ \Sigma \hookrightarrow \Sigma' = \Sigma' \hookrightarrow \Sigma' + \Sigma'' \circ \Sigma \hookrightarrow \Sigma'$ . Therefore, we have:

$$\begin{aligned} \text{Mod}(\Sigma \hookrightarrow \Sigma')(\text{Mod}(\alpha)(\text{Mod}(\Sigma'' \hookrightarrow \Sigma' + \Sigma'')(\mathcal{M}))) \\ = \\ \text{Mod}(\Sigma \hookrightarrow \Sigma')(\text{Mod}(\Sigma' \hookrightarrow \Sigma' + \Sigma'')(\mathcal{M})) \end{aligned}$$

As  $\Sigma' \setminus \Sigma$  is implicitly definable by  $T$ , we have:

$$\text{Mod}(\alpha)(\text{Mod}(\Sigma'' \hookrightarrow \Sigma' + \Sigma'')(\mathcal{M})) \equiv_{\Sigma'} \text{Mod}(\Sigma' \hookrightarrow \Sigma' + \Sigma'')(\mathcal{M})$$

We then have:

$$\text{Mod}(\alpha^{-1})(\text{Mod}(\alpha)(\text{Mod}(\Sigma'' \hookrightarrow \Sigma' + \Sigma'')(\mathcal{M}))) \models_{\Sigma''} \text{Sen}(\alpha)(T \cup E)$$

i.e.,  $\text{Mod}(\Sigma'' \hookrightarrow \Sigma' + \Sigma'')(\mathcal{M}) \models_{\Sigma''} \text{Sen}(\alpha)(T \cup E)$ . By the satisfaction condition, we conclude  $\mathcal{M} \models_{\Sigma' + \Sigma''} \text{Sen}(\alpha)(T \cup E)$ . □

**Theorem 4.15** *Let  $\mathcal{I}$  be an institution that satisfies CIP. Then,  $\mathcal{I}$  has the Beth property.*

**Proof** Let  $\Sigma' \setminus \Sigma$  be a signature that is implicitly definable for a  $\Sigma'$ -theory  $T$ . Let  $\varphi \in \text{Sen}(\Sigma' \setminus \Sigma)$ . Let  $\Sigma \hookrightarrow \Sigma''$  such that  $\Sigma' \approx \Sigma''$ . By Lemma 4.14, we have  $T \cup \{\varphi\} \models_{\Sigma' + \Sigma''} \text{Sen}(\alpha)(\varphi)$ . By CIP, there exists  $E' \subseteq \text{Sen}(\Sigma)$  such that  $T \cup \{\varphi\} \models_{\Sigma'} E'$  and  $E' \models_{\Sigma''} \text{Sen}(\alpha)(\varphi)$ . Let  $\mathcal{M} \in |\text{Mod}(\Sigma')|$  such that  $\mathcal{M} \models_{\Sigma'} E'$ . By the satisfaction condition,  $\text{Mod}(\alpha^{-1})(\mathcal{M}) \models_{\Sigma''} E'$ . Therefore,  $\text{Mod}(\alpha^{-1})(\mathcal{M}) \models_{\Sigma''} \text{Sen}(\alpha)(\varphi)$  whence by the satisfaction condition  $\mathcal{M} \models_{\Sigma'} \varphi$ . Hence,  $E' \models_{\Sigma'} \varphi$ , and then,  $T \cup E' \models_{\Sigma'} \varphi$ . □

**Example 4.16** Hence, all the examples of institutions given in Example 2.2 except **MFOL** have BP.<sup>10</sup> Indeed, signatures are restricted to inclusions and CIP only holds for (many-sorted) **FOL**, **LMFOL** and **LIMFOL** for commutative squares with one component injective on sorts, and only holds for (many-sorted) **HCL** and **EQL** for commutative squares with  $\sigma_2$  injective.

#### 4.4 Preservation of BP

In order to lighten the presentation, all the notions and results that will be defined and proved in this section will be stated for institution morphisms. However, they can easily be extended to institution comorphisms.

The preservation of BP through institution morphisms and comorphisms rests on the following condition:

**Definition 4.17 (Elementary equivalence preservation)** An institution morphism  $\mu : \mathcal{I} \rightarrow \mathcal{I}'$  is said to preserve elementary equivalence or has the elementary equivalence preservation property (EEPP) if and only if for every signature  $\Sigma$  in  $\mathcal{I}$  and for all models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  in  $Mod(\Sigma)$ , if  $\beta_\Sigma(\mathcal{M}_1) \equiv_{\Phi(\Sigma)} \beta_\Sigma(\mathcal{M}_2)$  then  $\mathcal{M}_1 \equiv_\Sigma \mathcal{M}_2$ .

Indeed, EEPP leads to the following result which is the first step to prove the preservation of BP through institution morphisms and comorphisms:

**Proposition 4.18** Let  $\mu : \mathcal{I} \rightarrow \mathcal{I}'$  be an institution morphism that has EEPP together with a functor  $\overline{\Phi} : Sig' \rightarrow Sig$  that preserves inclusion systems and satisfies  $\Phi \circ \overline{\Phi} = Id_{Sig'}$ . If  $\Sigma'_2 \setminus \Sigma'_1$  is implicitly definable by  $T'$  in  $\mathcal{I}'$  then  $\overline{\Phi}(\Sigma'_2) \setminus \overline{\Phi}(\Sigma'_1)$  is implicitly definable by  $\alpha_{\overline{\Phi}(\Sigma'_2)}(T')^\bullet$ .

**Proof** Let  $\mathcal{M}_1$  and  $\mathcal{M}_2$  be two models in  $Mod(\alpha_{\overline{\Phi}(\Sigma'_2)}(T')^\bullet)$  such that

$$Mod(\overline{\Phi}(\Sigma'_1) \hookrightarrow \overline{\Phi}(\Sigma'_2))(\mathcal{M}_1) = Mod(\overline{\Phi}(\Sigma'_1) \hookrightarrow \overline{\Phi}(\Sigma'_2))(\mathcal{M}_2)$$

By definition, we have

$$\beta_{\overline{\Phi}(\Sigma'_1)}(Mod(\overline{\Phi}(\Sigma'_1) \hookrightarrow \overline{\Phi}(\Sigma'_2))(\mathcal{M}_1)) = \beta_{\overline{\Phi}(\Sigma'_1)}(Mod(\overline{\Phi}(\Sigma'_1) \hookrightarrow \overline{\Phi}(\Sigma'_2))(\mathcal{M}_2))$$

By the naturality of  $\beta$ , we then deduce

$$Mod(\Sigma'_1 \hookrightarrow \Sigma'_2)(\beta_{\overline{\Phi}(\Sigma'_2)}(\mathcal{M}_1)) = Mod(\Sigma'_1 \hookrightarrow \Sigma'_2)(\beta_{\overline{\Phi}(\Sigma'_2)}(\mathcal{M}_2))$$

Moreover, by the satisfaction condition of institution morphisms,  $\beta_{\overline{\Phi}(\Sigma'_2)}(\mathcal{M}_1)$  and  $\beta_{\overline{\Phi}(\Sigma'_2)}(\mathcal{M}_2)$  are models of  $T'$ . Hence, by the hypothesis that  $\Sigma'_2 \setminus \Sigma'_1$  is implicitly definable by  $T'$ , we have:  $\beta_{\overline{\Phi}(\Sigma'_2)}(\mathcal{M}_1) \equiv_{\Sigma'_2} \beta_{\overline{\Phi}(\Sigma'_2)}(\mathcal{M}_2)$ . By EEPP, we then conclude  $\mathcal{M}_1 \equiv_{\overline{\Phi}(\Sigma'_2)} \mathcal{M}_2$ .  $\square$

<sup>10</sup>A counter-example that makes fail BP for **MFOL** can be found in [12].

Let us note that if we follow the standard definition of the implicit definability notion, *i.e.* imposing that the reduct functor  $Mod(\Sigma \hookrightarrow \Sigma') : Mod(T) \rightarrow Mod(\Sigma, \emptyset)$  in Definition 4.9 is injective, then EEPP comes to impose that  $\beta_\Sigma$  is injective.

EEPP is not sufficient to establish our preservation result. Actually, we need a stronger property that entails EEPP. This property is the following:

**Definition 4.19 (Conservative for sentences)** *Let  $\mu = (\Phi, \alpha, \beta) : \mathcal{I} \rightarrow \mathcal{I}'$  be an institution morphism together with a functor  $\bar{\Phi} : Sig' \rightarrow Sig$  that satisfies  $\Phi \circ \bar{\Phi} = Id_{Sig'}$ . It is conservative for sentences (CS) if and only if for every signature  $\Sigma'$  in  $\mathcal{I}'$ , for every  $\varphi \in Sen(\bar{\Phi}(\Sigma')) \setminus \alpha_{\bar{\Phi}(\Sigma')}(Sen'(\Sigma'))$ , there exists  $E' \subseteq Sen'(\Sigma')$  such that  $Mod(\varphi) = Mod(\alpha_{\bar{\Phi}(\Sigma')}(E'))$ .*

**Proposition 4.20** *If an institution morphism  $\mu$  is CS then it has EEPP.*

**Proof** Consider a signature  $\Sigma$  in  $\mathcal{I}$  and two models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  in  $Mod(\Sigma)$  such that  $\beta_\Sigma(\mathcal{M}_1) \equiv_{\Phi(\Sigma)} \beta_\Sigma(\mathcal{M}_2)$ . Let us show that  $\mathcal{M}_1 \equiv_\Sigma \mathcal{M}_2$ . Let  $\varphi \in Sen(\Sigma)$ . As  $\mu$  is CS, there exists  $E' \subseteq Sen'(\Phi(\Sigma))$  such that  $Mod(\{\varphi\}) = Mod(\alpha_\Sigma(E'))$ . Hence, we have the following equivalences:

$$\begin{aligned} \mathcal{M}_1 \models \varphi &\iff \mathcal{M}_1 \models_\Sigma \alpha_\Sigma(E') \\ &\iff \beta_\Sigma(\mathcal{M}_1) \models'_{\Phi(\Sigma)} E' \\ &\iff \beta_\Sigma(\mathcal{M}_2) \models'_{\Phi(\Sigma)} E' \\ &\iff \mathcal{M}_2 \models_\Sigma \varphi \end{aligned}$$

whence we conclude that  $\mathcal{M}_1 \equiv_\Sigma \mathcal{M}_2$ .  $\square$

A sufficient condition that entails CS is when  $\alpha_\Sigma$  is surjective. Hence, **RWL**  $\rightarrow$  **HCL<sub>T</sub>** is CS.

**Theorem 4.21** *Let  $\mu = (\Phi, \alpha, \beta) : \mathcal{I} \rightarrow \mathcal{I}'$  be a CS institution morphism together with a functor  $\bar{\Phi} : Sig' \rightarrow Sig$  that preserves inclusion systems and satisfies  $\Phi \circ \bar{\Phi} = Id_{Sig'}$ . If  $\mathcal{I}$  has BP then so does  $\mathcal{I}'$  as well.*

**Proof** Suppose that  $\Sigma'_2 \setminus \Sigma'_1$  is implicitly definable by a  $\Sigma'_2$ -theory  $T'$ . By Proposition 4.20 and Proposition 4.18, we have that  $\bar{\Phi}(\Sigma'_2) \setminus \bar{\Phi}(\Sigma'_1)$  is implicitly definable by  $\alpha_{\bar{\Phi}(\Sigma'_2)}(T')^\bullet$ . As  $\mathcal{I}$  has BP,  $\bar{\Phi}(\Sigma'_2) \setminus \bar{\Phi}(\Sigma'_1)$  is then explicitly definable relative to  $\alpha_{\bar{\Phi}(\Sigma'_2)}(T')^\bullet$ . Hence, for every  $\varphi \in Sen'(\Sigma'_2 \setminus \Sigma'_1)$  there exists  $E \subseteq Sen(\bar{\Phi}(\Sigma'_1))$  such that  $Mod(\alpha_{\bar{\Phi}(\Sigma'_2)}(T' \cup \{\varphi\})) = Mod(\alpha_{\bar{\Phi}(\Sigma'_1)}(T') \cup Mod(Sen(\bar{\Phi}(\Sigma'_1) \hookrightarrow \bar{\Phi}(\Sigma'_2))(E)))$ . As  $\mu$  is CS, there exists a set  $E' \subseteq Sen'(\Sigma'_1)$  such that  $Mod(E) = Mod(\alpha_{\bar{\Phi}(\Sigma'_1)}(E'))$ . Consequently, by the satisfaction property of institution morphisms, we have  $Mod'(T' \cup \{\varphi\}) = Mod'(T' \cup E')$ .  $\square$

**Corollary 4.22** *Let  $\mu = (\Phi, \alpha, \beta) : \mathcal{I} \rightarrow \mathcal{I}'$  be an institution morphism such that for every signature  $\Sigma$  in  $\mathcal{I}$ ,  $\alpha_\Sigma$  is surjective. Moreover,  $\mu$  is together with a functor  $\bar{\Phi} : Sig' \rightarrow Sig$  that preserves inclusion systems and satisfies  $\Phi \circ \bar{\Phi} = Id_{Sig'}$ . If  $\mathcal{I}$  has BP then so does  $\mathcal{I}'$  as well.*

## 4.5 The Beth property for all signature morphisms

Beth property for signature inclusions as previously defined comes from classical logic which does not consider non-injective signature morphisms. However, non-injective morphisms are very important in specification theory. Regarding the extension of the explicit definability notion, the Beth property can be simply generalised to any signature morphisms as follows:

**Definition 4.23 (Implicitly definable)** *Let  $\sigma : \Sigma \rightarrow \Sigma'$  be a signature morphism and let  $T$  be a  $\Sigma'$ -theory.  $\sigma$  is implicitly definable by  $T$  if and only if for all  $\mathcal{M}'$  and  $\mathcal{M}''$  in  $\text{Mod}(T)$ , if  $\text{Mod}(\sigma)(\mathcal{M}') = \text{Mod}(\sigma)(\mathcal{M}'')$ , then  $\mathcal{M}' \equiv_{\Sigma'} \mathcal{M}''$ .*

**Definition 4.24 (Explicitly definable)** *Let  $\sigma : \Sigma \rightarrow \Sigma'$  be a signature morphism and  $T$  be a  $\Sigma'$ -theory.  $\sigma$  is explicitly definable relative to  $T$  if for every  $\varphi \in \text{Sen}(\Sigma') \setminus \text{Sen}(\sigma)(\text{Sen}(\Sigma))$  there exists  $E \subseteq \text{Sen}(\Sigma)$  such that both following conditions are satisfied:*

1.  $T \cup \{\varphi\} \models_{\Sigma'} \text{Sen}(\sigma)(E)$
2.  $T \cup \text{Sen}(\sigma)(E) \models_{\Sigma'} \varphi$

Here, the Beth property establishes a correspondence between implicitly and explicitly definable properties. Indeed, we have the following result:

**Theorem 4.25** *Let  $\mathcal{I}$  be an institution that has signature pushouts and satisfies CIP. Let  $\sigma : \Sigma \rightarrow \Sigma'$  be a signature morphism and  $T$  be a  $\Sigma'$ -theory. Then,  $\sigma$  is implicitly definable for  $T$  if and only if  $\sigma$  is explicitly definable relative to  $T$ .*

**Proof**

( $\Rightarrow$ ) Let  $\sigma : \Sigma \rightarrow \Sigma'$  be a signature morphism which is implicitly definable for a  $\Sigma'$ -theory. Let  $\varphi \in \text{Sen}(\Sigma') \setminus \text{Sen}(\sigma)(\text{Sen}(\Sigma))$  be a  $\Sigma'$ -sentence. Let us consider the pushout of  $\sigma$  with itself:

$$\begin{array}{ccc} \Sigma & \xrightarrow{\sigma} & \Sigma' \\ \sigma \downarrow & & \downarrow \sigma_1 \\ \Sigma' & \xrightarrow{\sigma_2} & \Sigma'' \end{array}$$

Let us show that  $\text{Sen}(\sigma_1)(T) \cup \{\text{Sen}(\sigma_1)(\varphi)\} \models_{\Sigma''} \text{Sen}(\sigma_2)(\varphi)$ . Let us suppose that  $\mathcal{M} \models_{\Sigma''} \text{Sen}(\sigma_1)(T) \cup \{\text{Sen}(\sigma_1)(\varphi)\}$ . By Proposition 2.6, we have that  $\text{Mod}(\sigma_1)(\mathcal{M}) \models_{\Sigma'} T \cup \{\varphi\}$ . By the property of pushout, we have  $\text{Mod}(\sigma_2 \circ \sigma)(\mathcal{M}) = \text{Mod}(\sigma_1 \circ \sigma)(\mathcal{M})$ .

As  $\sigma$  is implicitly definable for  $T$ , we have that  $\text{Mod}(\sigma_2)(\mathcal{M}) \equiv_{\Sigma''} \text{Mod}(\sigma_1)(\mathcal{M})$ , whence we conclude that  $\text{Mod}(\sigma_2)(\mathcal{M}) \models_{\Sigma'} T \cup \{\varphi\}$ .

By the satisfaction condition, we have  $\mathcal{M} \models_{\Sigma''} \text{Sen}(\sigma_2)(T \cup \{\varphi\})$ , and then  $\mathcal{M} \models_{\Sigma''} \text{Sen}(\sigma_2)(\varphi)$ .

Hence, by CIP, there exists  $E' \subseteq \text{Sen}(\Sigma)$  such that  $T \cup \{\varphi\} \models_{\Sigma'} \text{Sen}(\sigma)(E')$  and  $\text{Sen}(\sigma)(E') \models_{\Sigma'} \varphi$ , whence we conclude  $T \cup \text{Sen}(\sigma)(E') \models_{\Sigma'} \varphi$ .

( $\Leftarrow$ ) Let  $\sigma : \Sigma \rightarrow \Sigma'$  be a signature morphism and two  $\Sigma'$ -models  $\mathcal{M}_1, \mathcal{M}_2 \in \text{Mod}(T)$  such that  $\text{Mod}(\sigma)(\mathcal{M}_1) = \text{Mod}(\sigma)(\mathcal{M}_2)$ . Let  $\varphi$  be a  $\Sigma'$ -sentence such that  $\mathcal{M}_i \models_{\Sigma'} \varphi$ . Let us show that  $\mathcal{M}_j \models_{\Sigma'} \varphi$  for  $i \neq j$ . By the satisfaction condition of institution, this is obvious whenever  $\varphi \in \text{Sen}(\sigma)(\Sigma)$ . Therefore, let us suppose that  $\varphi \in \text{Sen}(\Sigma') \setminus \text{Sen}(\sigma)(\Sigma)$ . As  $\sigma$  is explicitly definable relative to  $T$ , there exists  $E \subseteq \text{Sen}(\Sigma)$  such that  $T \cup \text{Sen}(\sigma)(E) \models_{\Sigma'} \varphi$  and  $T \cup \{\varphi\} \models_{\Sigma'} E$ . Let us suppose that  $\mathcal{M}_j \not\models_{\Sigma'} \varphi$ . This means that there is some  $\psi \in E$  such that  $\mathcal{M}_j \not\models_{\Sigma'} \text{Sen}(\sigma)(\psi)$ . By the satisfaction condition, we then have that  $\text{Mod}(\sigma)(\mathcal{M}_j) \not\models_{\Sigma} \psi$  and then  $\text{Mod}(\sigma)(\mathcal{M}_i) \not\models_{\Sigma} \psi$ . By the satisfaction condition, we then conclude that  $\mathcal{M}_i \not\models_{\Sigma} \text{Sen}(\sigma)(\psi)$ . But, by  $T \cup \{\varphi\} \models_{\Sigma'} \text{Sen}(\sigma)(E)$  and the satisfaction condition, we also have that  $\mathcal{M}_i \models_{\Sigma'} \text{Sen}(\sigma)(\psi)$  what is a contradiction.  $\square$

**Example 4.26** *By Example 3.2, PL and the unsorted sub-institutions of FOL, LMFOL and LIMFOL have BP for every signature morphism  $\sigma : \Sigma \rightarrow \Sigma'$  and every  $\Sigma'$ -theory  $T$ . On the contrary, (many-sorted) FOL, LMFOL and LIMFOL have BP only for signature morphisms injective on sorts, and HCL and EQL have BP for injective (on sorts and operations) signature morphisms.*

## 5 Conclusion

We have formulated and proven a general form of the Beth definability theorem in the framework of institutions. This result has been proven by following the approach applied in conventional model theory, that is showing that the Beth definability property is a consequence of the Craig interpolation property.

The Beth definability theorem has first been proven in dependence of the categorical axiomatization of set-theoretical operations given in [10] in order to abstract the notion of signature symbols. Then, this result has been generalised to any signature morphisms. Thus, the classic definability problem of defining a new symbol  $\chi$  with respect to a given signature  $\Sigma$  (defining a signature inclusion  $\Sigma \hookrightarrow \Sigma \cup \{\chi\}$ ) is generalised to any kind of signature morphism. Finally, we have presented the preservation of BP through institution morphism and comorphisms satisfying certain specific properties which allows the transfer of the Beth definability theorem from one institution to another.

### Acknowledgements

Special thanks are due to Razvan Diaconescu and Till Mossakowski for many fruitful and instructive discussions and numerous useful comments. The authors also especially thank José Luiz Fiadeiro for its careful reading of this text.

## References

- [1] M. Aiguier, C. Gaston, and P. Le Gall. Feature Specification: a Logic-Independent Approach. Technical report, Université d'Évry, 2004.



- [2] M. Aiguier and P.-Y. Schobbens. A Note on Robinson Consistency Lemma. Technical report, Université d'Évry, 2006. Available at [www.ibisc.univ-evry.fr/~aiguier](http://www.ibisc.univ-evry.fr/~aiguier).
- [3] J Barwise. Axioms for Abstract Model Theory. *Annals of Mathematical Logic*, 7:221–265, 1974.
- [4] T. Borzyszkowski. Generalized Interpolation in CASL. *Information Processing letters*, 2001.
- [5] T. Borzyszkowski. Logical Systems for Structured Specifications. *Theoretical Computer Science*, 286:197–245, 2002.
- [6] R. Diaconescu. Herbrand Theorems in Arbitrary Institutions. *Information Processing Letters*, 90(1):29–39, 2002.
- [7] R. Diaconescu. Institution-Independent Ultraproducts. *Fundamenta Informaticae*, 55(3-4):321–348, 2003.
- [8] R. Diaconescu. An Institution-Independent Proof of Craig Interpolation Property. *Studia Logica*, 77(1):59–79, 2004.
- [9] R. Diaconescu. Interpolation in Grothendieck Institutions. *Theoretical Computer Science*, 55(3-4):321–348, 2004.
- [10] R. Diaconescu, J.A. Goguen, and P. Stefaneas. Logical Support for Modularisation. In G. Huet and G. Plotkin, editors, *Logical Environment*, pages 83–130, 1993.
- [11] T. Dimitrakos and T. Maibaum. On a Generalized Modularization Theorem. *Information Processing Letter*, 74:65–71, 2000.
- [12] K. Fine. Failures for the Interpolation Lemma in Quantified Modal Logic. *Journal of Symbolic Logic*, 44:201–206, 1979.
- [13] D. Gabbay and L. Maksimova. *Interpolation and Definability Volume 1*. Clarendon Press - Oxford, 2004.
- [14] J. Goguen and R. Burstall. A Study in the Foundation of Programming Methodology: Specifications, Institutions, Charters and Parchments. In D Pitt, S. Abramsky, A. Poigné, and D. Rydeheard, editors, *Proceedings of the Conference on Category Theory and Computer Programming*, volume 240 of *Lecture Notes in Computer Science*, pages 313–333. Springer Verlag, 1985.
- [15] J. Goguen and R. Burstall. Institutions: Abstract Model Theory for Specification and Programming. *Journal of the Association for Computing Machinery*, 39(1):95–146, 1992.
- [16] D. Găină and A. Popescu. An Institution-Independent Proof of Robinson Consistency Theorem. *Studia Logica*, 2006. To appear.

- [17] A. Salibra and G. Scollo. Compactness and Löwenheim-Skolem Properties in Pre-Institution Categories. In C.Rauszer, editor, *Algebraic Methods in Logic and in Computer Science*, volume 28, pages 67–94. Banach Center Publications, 1993.
- [18] A. Salibra and G. Scollo. Interpolation and Compactness in Categories of Pre-Institutions. *Mathematical Structures in Computer Science*, 6(3):261–286, 1996.
- [19] A. Sernadas, C. Sernadas, and C. Caleiro. Fibring of Logics as a Categorical Construction. *Journal of Logic and Computation*, 9(2):149–179, 1999.
- [20] A. Tarlecki. On the Existence of Free Models in Abstract Algebraic Institutions. *Theoretical Computer Science*, 37(3):269–304, 1985.
- [21] A. Tarlecki. Quasi-Varieties in Abstract Algebraic Institutions. *Journal of Computer and System Science*, 33(3):269–304, 1986.
- [22] A. Tarlecki. *Algebraic Foundations of Systems Specification*, chapter Institutions: An Abstract Framework for Formal Specifications. IFIP State-of-the-Art Reports. Springer Verlag, 1999.
- [23] A. Tarski. The Semantic Conception of Truth. *Philos. Phenomenological Research*, 4:13–47, 1944.
- [24] P. Veloso and S. Fiadeiro, J.-L.Veloso. On Local Modularity and Interpolation in Entailment Systems. *Information Processing Letter*, 82:203–211, 2002.
- [25] P. Veloso and T. Maibaum. On the Modularization Theorem for Logical Specifications. *Information Processing Letter*, 53(5):287–293, 1995.