

# Institutionalising Ontology-Based Semantic Integration

Marco Schorlemmer<sup>a</sup>, Yannis Kalfoglou<sup>b</sup>

<sup>a</sup> *IIIA – Artificial Intelligence Research Institute, CSIC, Catalonia, Spain*

<sup>b</sup> *School of Electronics and Computer Science, University of Southampton, UK*

**Abstract.** We address what is still a scarcity of general mathematical foundations for ontology-based semantic integration underlying current knowledge engineering methodologies in decentralised and distributed environments. After recalling the first-order ontology-based approach to semantic integration and a formalisation of ontological commitment, we propose a general theory that uses a syntax- and interpretation-independent formulation of language, ontology, and ontological commitment in terms of institutions. We claim that our formalisation generalises the intuitive notion of ontology-based semantic integration while retaining its basic insight, and we apply it for eliciting and hence comparing various increasingly complex notions of semantic integration and ontological commitment based on differing understandings of semantics.

Keywords: semantic integration, ontologies, ontological commitment, institutions

## 1. Introduction

Knowledge engineers have long realised that, for two or more information systems (databases, agents, peers, software components, expert systems, etc.) to be both syntactically and semantically interoperable, they will need to commit to a shared conceptualisation of the application domain. Commonly, this is achieved by producing an explicit specification of this conceptualisation — what has become to be known in computer science as an *ontology* — and by defining each system’s local language in terms of the ontology’s vocabulary. This sort of integration is dubbed “semantic” precisely because it assumes that the ontology consists of some sort of structured representation  $O$  — coming thus equipped with a precise semantics for the structure it holds — and because each system’s local language is *interpreted* in  $O$  (for instance, in the technical sense of a theory interpretation as defined in (Enderton, 2002), when  $O$  is a theory in first-order logic).

### 1.1. Ontologies and the Semantic Web

Because ontologies have been advocated as a way to make a shared conceptualisation explicit, thus enabling two systems to share the same ontological commitment, they have drawn much attention and been thoroughly exploited for knowledge sharing and semantic integration. But before ontologies became popular, knowledge engineers hardly ever had to work with more than one ontology at a time. Even in cases where multiple ontologies were used (see, e.g., (Borst et al., 1997)), these were mostly controlled experiments (e.g., (Uschold et al., 1998)) in moderated environments (such as (Farquhar et al., 1997)). Nowadays, however, the practice is somewhat different. Modern trends in knowledge management dictate that we should expect to work more and more within distributed and open-ended environments like the Web. That fact alone has had a significant impact on knowledge representation with ontologies.

Firstly, we observe that sourcing ontologies is far easier today than it was in the recent past. Once Semantic Web technologies became more mature (like, for example, the RDF language (Lassila and Swick, 1999) or the OWL family of languages (McGuinness and van Harmelen, 2004) backed by the World Wide Web Consortium (W3C)), a plethora of ontologies were made readily available and accessible via

the Web.<sup>1</sup> Even if the quality or the purpose served by these ontologies is questionable from a strict knowledge-representation point of view, their impact on practice is undisputed.

Second, due to the nature of the environment that most ontologies operate in (the Semantic Web, for example), it is more likely that we will need more than one ontology to achieve knowledge sharing. It is increasingly unlikely that a single ontology will both adequately capture the domain in question and also be consensual among all interested parties.

Third, strict knowledge engineering practice is difficult to enforce when dealing with outsourced ontologies. Syntactic compliance with Semantic Web standards (like OWL) is not enough to guarantee that our inferences will make sense and that automated reasoning will be possible. It is not uncommon to find subtle differences in meaning between any two ontologies even if they represent the same domain and are encoded in the same formalism.

Fourth, there are a number of reasons that go beyond computational reasoning: (a) *social factors*, like for example the impact that ontology-based codification of knowledge can have in a real world environment (e.g., does it facilitate or complicate human to human knowledge sharing?); (b) *contextual reasoning* (how and when should an ontology take into account or represent contextual information?); (c) *social agreements* (are ontologies too formal for enforcing or facilitating social agreements between agents —human or artificial?). The answers to these questions are not easy to find neither are they clearly understood.

## 1.2. Scarcity of Mathematical Foundations

Recently, several scientific events (Doan et al., 2004a; Kalfoglou et al., 2005; Benjamins et al., 2006) and journal issues (Doan et al., 2004b; Noy et al., 2005; Shvaiko and Euzenat, 2007) have been solely devoted to the topic of ontology-based semantic integration, and there exist also several comprehensive surveys of the field, covering database schema matching (Rahm and Bernstein, 2001), information integration (Wache et al., 2001), ontology mapping (Ding and Foo, 2002; Kalfoglou and Schorlemmer, 2003b; Shvaiko and Euzenat, 2005), and semantic integration (Noy, 2004). However, what we have frequently observed in the works reported in these events and surveys is that the mathematical foundations underlying most research into the problem of semantic heterogeneity have so far been only marginally approached (Kalfoglou et al., 2004). Often solutions are presented without precise definitions of the concepts at work, such as ‘language’, ‘ontology’, ‘semantics’, ‘ontological commitment’, or ‘semantic integration.’ This often brings forth an overly optimistic view of the capabilities of ontology-based technology for supporting large-scale system interoperability.

There exist some notable exceptions, though: Guarino, Carrara and Giaretta, for instance, proposed a mathematical model for ontological commitment (Guarino et al., 1994; Guarino, 1998a); Bench-Capon, Malcolm and Shave formalised ontology and their compatibility (Bench-Capon and Malcolm, 1999; Bench-Capon et al., 2003); and Ciocoiu and Nau provided a formal definition of ontology-based semantics and translation (Ciocoiu and Nau, 2000). More recently, Menzel (Menzel, 2002, 2005) and Grüninger (Grüninger, 2005) have also proposed basic theories of ontology and semantic integration. The common leitmotif of all these proposals is the use of some logical system (first-order logic, order-sorted equational logic, modal logic, etc.) to formalise the idea that two systems are semantically interoperable if, after data is transmitted from a sender system to a receiver, all implications made by one system hold and are provable by the other, and that there is a logical equivalence between those implications.

Still, most semantic integration technology focuses on mapping at the terminological level, while taking the unstated assumption that the mechanism by which terminology is interpreted, is actually the same across different communities. Current mapping techniques map ontology entities such as concept and role descriptions using sophisticated logical inferences; but these inferences nevertheless are based on the strong assumption of a common interpretation mechanism, such as a first-order model-theoretic semantics, or by assuming a shared universe of interpretation.

<sup>1</sup>By using, for example, Swoogle (Ding et al., 2004).

### 1.3. An Interpretation-Independent Approach

A careful look at the several formal approaches to semantic integration mentioned above reveals many different understandings of semantics depending on the interoperability scenario under consideration. Consequently, by choosing a particular logical system for the sake of formalising ontology-based semantic integration one also commits to a particular understanding of semantics. Therefore, what we need in order to successfully tackle the mathematical foundations of semantic integration is not so much a framework that establishes one particular semantic perspective (model-theoretic, property-theoretic, proof-theoretic, instance-based, etc.), but instead we need a framework that successfully captures semantic integration despite the different treatments of semantics.

We believe that in order to address the semantic heterogeneity problem in all its complexity, we need to take into account locality and difference not only at the terminological level, but also at the level of the interpretation mechanisms of a community, and the actual scope and use of ontologies by means of particular communities. This requires an adequate mathematical framework that is general enough to cope with heterogeneity both in terminology *and* in interpretation mechanism, but that is also concrete enough for providing insights into the actual deployment of semantic integration technology. We resort for this reason to the extensive work that has been carried out in the field of formal software specification using *institutions* (Goguen and Burstall, 1992) and propose an interpretation-independent characterisation of ontology commitment.

It should be noted, though, that in this paper we do not want to wander into a philosophical discussion on the notions of ‘ontology’ and ‘ontological commitment.’ Others provide an in-depth analysis of these notions (see, e.g., (Guarino, 1995; Smith, 2003)). Rather, we will build upon the understanding of these notions by the artificial intelligence and information systems communities. For this reason we shall first recall the first-order, model-theoretic approach to ontology-based semantic integration, and then further extend it with a formalisation of ontological commitment due to Guarino, Carrara and Giaretta (Guarino et al., 1994; Guarino, 1998a). This will help us introduce the technical apparatus needed to lift these notions into the institutional framework. In this sense, this article aims at contributing towards a “general theory of ontology translation and integration” as advocated by Goguen in (Goguen, 2005), but focusing on an institutionalised conceptualisation of ontological commitment that is compatible with Guarino et al.’s basic insights.

We have already advocated elsewhere for a formal, general foundation for semantic integration (Schorlemmer and Kalfoglou, 2003; Kalfoglou and Schorlemmer, 2004) demonstrating its applicability to various scenarios, such as ontology mapping (Kalfoglou and Schorlemmer, 2003a), ontology coordination (Schorlemmer and Kalfoglou, 2005), ontology-alignment interaction models (Schorlemmer et al., 2007), and situated semantic alignment (Atencia and Schorlemmer, 2007), although in neither of them we explored the interpretation-independence that institutions provide.

## 2. Ontology-Based Semantic Integration

We shall initially be concerned with semantic integration understood as the integration of two information systems by virtue of the interpretation of their respective knowledge representation vocabularies into a reference theory—an ontology—expressible in the language of first-order logic. In practice, semantic integration is often carried out on subsets of first-order logic, such as description logics (DL), for which reasoning has good computational properties. This is, for instance, the approach followed by Calvanese and De Giacomo in their ontology integration system for database schemata (Calvanese and De Giacomo, 2005); W3C, too, has embraced DLs in order to develop the OWL recommendation for ontology representation (McGuinness and van Harmelen, 2004). Another example is the focus of Giunchiglia, Marchese and Zaihrayeu on propositional DLs in order to use fast SAT provers for matching taxonomically organised vocabularies (Giunchiglia et al., 2006). In contrast, the Process Specification Language (PSL) is an

example of a semantic integration initiative based on full first-order logic that uses invariants to define interpretations of local vocabulary into PSL (Grüninger and Kopena, 2005).

By *vocabulary* we mean a collection  $\Sigma$  of words and symbols used by an information system to represent and organise its local knowledge. In a formal, logic-based representation language the vocabulary is constituted by the non-logical symbols used to form sentences and formulae and usually referred to as *parameters* or *signature*. The *language* is then the set of all syntactically well-formed formulae over a given vocabulary  $\Sigma$ , which we shall denote with  $L = \text{Sen}(\Sigma)$ . We call the elements of  $\text{Sen}(\Sigma)$ , *sentences*.

In declarative representation languages, knowledge is represented and organised by means of theories. We call a *theory* a pair  $T = (\Sigma, \Gamma)$ , where  $\Gamma \subseteq \text{Sen}(\Sigma)$ . The sentences in  $\Gamma$  are the *axioms* of the theory. Finally, in order to capture the relationship between theories, we call a *theory interpretation* a map between the underlying languages of theories that respects theoremhood of axioms. That is, a function  $\alpha : \text{Sen}(\Sigma) \rightarrow \text{Sen}(\Sigma')$  is also a theory interpretation  $\alpha : T \rightarrow T'$  between theories  $T = (\Sigma, \Gamma)$  and  $T' = (\Sigma', \Gamma')$  if, and only if, for all  $\varphi \in \text{Sen}(\Sigma)$ , if  $\varphi$  is a logical consequence of  $\Gamma$  then  $\alpha(\varphi)$  is a logical consequence of  $\Gamma'$ , i.e.,  $\Gamma \models \varphi$  implies  $\Gamma' \models \alpha(\varphi)$ . It is a *faithful* theory interpretation if, in addition,  $\Gamma' \models \alpha(\varphi)$  implies  $\Gamma \models \varphi$ .

Recall that, in first-order logic, theory interpretations are usually defined by first giving a translation  $\pi$  of vocabulary symbols (and the symbol  $\forall$ ) to first-order formulae as follows (see (Enderton, 2002)):<sup>2</sup>

- $\pi$  assigns to  $\forall$  a formula  $\pi_{\forall}$  over  $\Sigma'$  in which at most one variable  $x$  occurs free;
- $\pi$  assigns to each  $n$ -ary relation symbol  $r \in \Sigma$  a formula  $\pi_r$  over  $\Sigma'$  in which at most  $n$  variables occur free;

and then extending  $\pi$  recursively to a translation of formulae to formulae: if  $\varphi$  is an atomic formula with  $n$ -ary relation symbol  $r$ ,  $\pi_{\varphi} = \pi_r$  applied to the same set of variables and constants; otherwise  $\pi_{\neg\varphi} = \neg\pi_{\varphi}$ ,  $\pi_{\varphi \rightarrow \psi} = \pi_{\varphi} \rightarrow \pi_{\psi}$ , and  $\pi_{\forall x \varphi} = \forall x (\pi_{\forall} \rightarrow \pi_{\varphi})$ . A theory interpretation  $\alpha : (\Sigma, \Gamma) \rightarrow (\Sigma', \Gamma')$  is then the restriction of  $\pi$  to sentences of  $\text{Sen}(\Sigma)$ , but only if also  $\Gamma' \models \exists x \pi_{\forall}$ .

By the way a theory interpretation  $\alpha : T \rightarrow T'$  is defined for first-order logic, one can extract from each first-order  $\Sigma'$ -structure  $\mathfrak{S}$  a first-order  $\Sigma$ -structure  $\beta(\mathfrak{S})$  as follows: let the domain of  $\beta(\mathfrak{S})$  be the subset of elements of the domain of  $\mathfrak{S}$  determined by the formula  $\pi_{\forall}$ , and let  $\beta(\mathfrak{S})$  assign to each relation symbol  $r \in \Sigma$  the relation determined by the formula  $\pi_r$  in  $\mathfrak{S}$ . In addition, for all  $\varphi \in \text{Sen}(\Sigma)$  and first-order  $\Sigma'$ -structure  $\mathfrak{S}$ , we have that  $\mathfrak{S}$  satisfies  $\alpha(\varphi)$  if, and only if,  $\beta(\mathfrak{S})$  satisfies  $\varphi$ . Consequently, if we write  $\text{Mod}(T)$  for the set of all  $\Sigma$ -structures satisfying the axioms of  $T$  and  $\text{Mod}(T')$  for the set of all  $\Sigma'$ -structures satisfying the axioms of  $T'$ ,  $\beta$  restricts to a function mapping structures in  $\text{Mod}(T')$  to structures in  $\text{Mod}(T)$  and is called a *structure reduct*.

**Definition 1 (First-Order Semantic Integration)** *We say that two theories  $T_1$  and  $T_2$  are semantically integrated with respect to  $T$ , if*

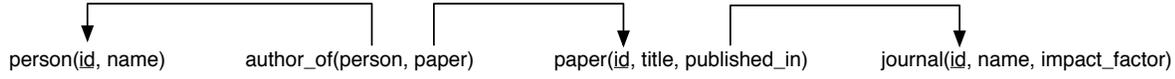
- *there exist theory interpretations  $\alpha_1 : T_1 \rightarrow T$  and  $\alpha_2 : T_2 \rightarrow T$ ;*
- *there exist structure reducts  $\beta_1 : \text{Mod}(T) \rightarrow \text{Mod}(T_1)$  and  $\beta_2 : \text{Mod}(T) \rightarrow \text{Mod}(T_2)$ ;*
- *$\text{Mod}(T) \neq \emptyset$ .*

*We call  $\mathfrak{I} = \{\alpha_i : T_i \rightarrow T; \beta_i : \text{Mod}(T) \rightarrow \text{Mod}(T_i)\}_{i=1,2}$  a semantic integration of local theories  $T_1$  and  $T_2$  with respect to reference theory  $T$ . Two languages  $L_1 = \text{Sen}(\Sigma_1)$  and  $L_2 = \text{Sen}(\Sigma_2)$  are semantically integrated with respect to  $T$  if the theories  $(\Sigma_1, \emptyset)$  and  $(\Sigma_2, \emptyset)$  are.*

We shall call  $\mathfrak{I}$  a *faithful* semantic integration if its theory interpretations are faithful. The structure reducts  $\beta_i : \text{Mod}(T) \rightarrow \text{Mod}(T_i)$  restrict the set of models for language  $\text{Sen}(\Sigma_i)$  to those  $\Sigma_i$ -structures that lie in the image of  $\beta_i$  of  $T$ -models, i.e., to those  $\Sigma_i$ -structures that are compatible with the models of the reference theory  $T$ . Consequently, we can say that  $\alpha_i$  and  $\beta_i$  capture together the commitment of  $\text{Sen}(\Sigma_i)$  to a conceptualisation —to the models of reference theory  $T$ .

<sup>2</sup>This amounts to what has been called a *derived signature morphism* in the literature.

Relational schema of information system  $A$ :



DL T-Box of information system  $B$ :

$$\begin{aligned} \text{Researcher} &\sqsubseteq \exists \text{name}.\top \\ \text{Article} &\sqsubseteq \exists \text{author}.\top \sqcap \exists \text{title}.\top \\ \text{Journal} &\sqsubseteq \exists \text{name}.\top \sqcap \exists \text{hasArticle}.\top \sqcap \exists \text{impactFactor}.\top \end{aligned}$$

Reference theory (or ontology)  $T$  of the scientific publications domain (using a fragment of the AKT Reference Ontology available at <http://www.aktors.org/publications/ontology/>):

$$\begin{aligned} \forall x ( \text{Working\_Person}(x) &\rightarrow ( \text{Tangible\_Thing}(x) \wedge \exists y ( \text{String}(y) \wedge \text{Name}(x, y) ) ) ) \\ \forall x ( \text{Researcher}(x) &\rightarrow \text{Working\_Person}(x) ) \\ \forall x ( \text{Composite\_Publication}(x) &\rightarrow ( \text{Tangible\_Thing}(x) \wedge \exists y ( \text{String}(y) \wedge \text{Name}(x, y) ) \\ &\quad \wedge \exists z ( \text{Publication}(z) \wedge \text{Has\_Publication}(x, z) ) ) ) ) \\ \forall x ( \text{Journal}(x) &\rightarrow ( \text{Composite\_Publication}(x) \wedge \exists y ( \text{Article}(y) \wedge \text{Has\_Article}(x, y) ) \\ &\quad \wedge \exists z ( \text{Real}(z) \wedge \text{Impact\_Factor}(x, z) ) ) ) ) \\ \forall x \forall y ( \text{Has\_Article}(x, y) &\rightarrow \text{Has\_Publication}(x, y) ) \\ \forall x ( \text{Proceedings}(x) &\rightarrow ( \text{Composite\_Publication}(x) \wedge \exists y ( \text{Paper}(y) \wedge \text{Has\_Paper}(x, y) ) ) ) ) \\ \forall x \forall y ( \text{Has\_Paper}(x, y) &\rightarrow \text{Has\_Publication}(x, y) ) \\ \forall x ( \text{Publication}(x) &\rightarrow ( \text{Tangible\_Thing}(x) \wedge \exists y ( \text{Researcher}(y) \wedge \text{Author}(x, y) ) \wedge \exists z ( \text{String}(z) \wedge \text{Title}(x, z) ) ) ) ) \\ \forall x ( \text{Article}(x) &\rightarrow \text{Publication}(x) ) \\ \forall x ( \text{Paper}(x) &\rightarrow \text{Publication}(x) ) \end{aligned}$$

Maps  $\alpha_A$  and  $\alpha_B$  of sentences, defined over the recursive structure of sentences of the first-order languages of the relational schema and the DL T-Box as follows:

$$\begin{aligned} \alpha_A(\text{person}(p, n)) &= \text{Researcher}(p) \wedge \text{String}(n) \wedge \text{Name}(p, n) \\ \alpha_A(\text{author\_of}(p, a)) &= \text{Researcher}(p) \wedge \text{Article}(a) \wedge \text{Author}(a, p) \wedge \exists j ( \text{Journal}(j) \wedge \text{Has\_Article}(j, a) ) \\ \alpha_A(\text{paper}(a, t, j)) &= \text{Article}(a) \wedge \text{String}(t) \wedge \text{Journal}(j) \wedge \text{Has\_Article}(j, a) \wedge \text{Title}(a, t) \\ \alpha_A(\text{journal}(j, n, f)) &= \text{Journal}(j) \wedge \text{String}(n) \wedge \text{Real}(f) \wedge \text{Name}(j, n) \wedge \text{Impact\_Factor}(j, f) \\ \alpha_B(\text{Article}(x)) &= \text{Publication}(x) \end{aligned}$$

Map  $\alpha_B$  is the identity on the remaining parameters of  $B$ 's language. It is easy to prove that  $\alpha_A$  and  $\alpha_B$  are indeed theory interpretations.

Fig. 1. Example of a semantic integration of system  $A$  based on a relational schema and system  $B$  based on a DL T-Box with respect to reference theory  $T$ .

Note that the above definition also comprises *trivial* integrations, where  $T$  is the theory over the disjoint union of local signatures, with the disjoint union of local axioms; and it makes theories with contradicting theorems impossible to integrate. The first problem can be solved by adding additional conditions to the definition of semantic integration, while the second problem can be tackled by integrating subtheories of the original theories. In this paper, though, we choose to stay with the above basic definition of semantic integration and leave further refinements of our framework for the future.

Figure 1 shows an example of the semantic integration of two systems that use varying data models and vocabulary: system  $A$  based on a relational schema and system  $B$  based on a DL T-Box. The semantic integration is realised with respect to a reference theory  $T$  —an ontology specifying a conceptualisation of the scientific publications domain. Both relational schema and DL T-Box are notational variants of fragments of first-order logic. Semantic integration can therefore be formalised by means of first-order

---

A SQL query over  $A$ 's relational schema:

```

SELECT person.name
FROM   person, author_of
WHERE  author_of.person = person.id
       AND NOT person.name = "Ann Smith"
       AND author_of.paper IN (SELECT paper.id
                               FROM   person, author_of, paper, journal
                               WHERE  author_of.person = person.id
                                     AND author_of.paper = paper.id
                                     AND paper.published_in = journal.id
                                     AND person.name = "Ann Smith"
                                     AND journal.impact_factor > 1.0)

```

The SQL query as sentence  $\varphi$  in first-order logic (actually, first-order logic with built-in number comparison):

$$\begin{aligned}
& \exists n_1 \exists p_1 \exists a \exists p_2 \exists n_2 \exists t \exists j \exists n \exists f \\
& (person(p_1, n_1) \wedge n_1 \neq \text{"Ann Smith"} \wedge author\_of(p_1, a) \\
& \wedge person(p_2, n_2) \wedge n_2 = \text{"Ann Smith"} \wedge author\_of(p_2, a) \\
& \wedge paper(a, t, j) \wedge journal(j, n, f) \wedge f > 1.0)
\end{aligned}$$

The interpretation  $\alpha_A(\varphi)$  of the query in  $T$ :

$$\begin{aligned}
& \exists n_1 \exists p_1 \exists a \exists p_2 \exists n_2 \exists t \exists j \exists n \exists f \\
& (Researcher(p_1) \wedge String(n_1) \wedge Name(p_1, n_1) \wedge n_1 \neq \text{"Ann Smith"} \wedge Researcher(p_1) \wedge Article(a) \wedge Author(a, p_1) \\
& \wedge Researcher(p_2) \wedge String(n_2) \wedge Name(p_2, n_2) \wedge n_2 = \text{"Ann Smith"} \wedge Researcher(p_2) \wedge Article(a) \wedge Author(a, p_2) \\
& \wedge Article(a) \wedge String(t) \wedge Journal(j) \wedge Has\_Article(j, a) \wedge Title(a, t) \\
& \wedge Journal(j) \wedge String(n) \wedge Real(f) \wedge Name(j, n) \wedge Impact\_Factor(j, f) \wedge f > 1.0)
\end{aligned}$$

Fig. 2. Interpretation of a SQL query of information system  $A$  into ontology  $T$ .

---

theory interpretations. Figure 2 shows the interpretation of a SQL query formulated according to  $A$ 's relational schema of Figure 1 in the ontology  $T$  of the scientific publications domain.

At the core of any formal approach to ontology-based semantic integration lies the assumption that interoperability should be formalised in terms of logical consequence. This is so because, by virtue of two local languages  $L_1 = Sen(\Sigma_1)$  and  $L_2 = Sen(\Sigma_2)$  committing to the same conceptualisation specified by  $T$ , we can check if a sentence  $\varphi \in L_1$  follows from a set of sentences  $\Gamma \subseteq L_2$  by checking if its interpretation into  $T$  is a logical consequence of both  $T$  and the interpretation of  $\Gamma$ 's sentences into  $T$ . This interoperability can be formally expressed by defining a consequence relation directly between sets of sentences of  $Sen(\Sigma_1)$  and sentences of  $Sen(\Sigma_2)$ :

**Definition 2 (First-Order Ontology-Based Consequence)** *Let  $\mathfrak{J} = \{\alpha_i : T_i \rightarrow T; \beta_i : Mod(T) \rightarrow Mod(T_i)\}_{i=1,2}$  be a semantic integration of  $T_1 = (\Sigma_1, \Gamma_1)$  and  $T_2 = (\Sigma_2, \Gamma_2)$  with respect to  $T$ , and let  $\varphi \in Sen(\Sigma_1)$  and  $\Gamma \subseteq Sen(\Sigma_2)$ . We say that  $\varphi$  is an ontology-based consequence of  $\Gamma$ , written  $\Gamma \models_{\mathfrak{J}} \varphi$ , if  $T \cup \alpha_2(\Gamma) \models \alpha_1(\varphi)$ .*

First-order ontology-based consequence corresponds to what Ciocoiu and Nau have called *ontology-based partial translation* (Ciocoiu and Nau, 2000). Figure 3 summarises the syntactic and semantic links that are established in the ontology-based approach to semantic integration.

### 3. Formalising Ontological Commitment

In the previous section we have recalled the theoretical basis underlying the first-order approach to ontology-based semantic integration, namely through *first-order interpretations of separate vocabularies*

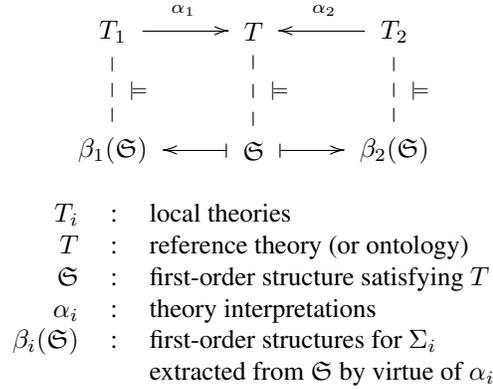


Fig. 3. First-order ontology-based semantic integration.

into a common ontology. But although this is currently the dominant approach to provide a formal foundation to semantic integration, not everyone agrees with the role of first-order logic and model-theoretic semantics to formalise ontological commitment. Guarino, Carrara and Giaretta, for instance, support the view that first-order logic is not expressive enough for specifying conceptualisations, and hence for writing down ontologies, because it cannot capture the intensionality that comes with the conceptualisation process of a fragment of the world. They argue that in order to specify a shared conceptualisation one needs to enrich a first-order language with modalities (Guarino et al., 1994; Guarino and Giaretta, 1995; Guarino, 1998a), otherwise one gets a very rough characterisation of the ontological commitment. Hence, in their view, an ontology is a theory in an expressive, but computationally inefficient logical language. Similar in spirit is Menzel’s proposal for an ontology theory that uses a property-theoretic approach, drawing from computational linguistics, and where entailment and equivalence are not model-theoretically defined, but axiomatised in a logical language for ontology theory (Menzel, 2002).

### 3.1. An Intensional Semantics

Taking Gruber’s highly cited definition of an ontology as an “explicit specification of a conceptualisation” (Gruber, 1993), Guarino et al. reflect on the appropriate mathematical structure accounting for a conceptualisation, questioning Genesereth and Nilsson’s use of extensional relations for that purpose. An extensional relation of arity  $n$  over domain of discourse  $D$  is a subset  $\rho \subseteq D \times \dots \times D$  (i.e.,  $\rho \in 2^{(D^n)}$ ). It reflects the relationship in which elements of the domain of discourse stand in a particular state of affairs. A conceptualisation, Guarino et al. argue, should account for the *meaning* of the relation instead, which cannot coincide with the extension of this relation in one particular state of affairs. As an alternative, Guarino et al. suggest to use intensional relations. An intensional relation of arity  $n$  over domain of discourse  $D$  is a function  $\rho : W \rightarrow 2^{(D^n)}$  (i.e.,  $\rho \in (2^{(D^n)})^W$ ), where  $W$  is a non-empty set of *possible worlds* or *states of affairs*. For a particular state of affairs  $w$ ,  $\rho(w)$  is the extensional relation in that particular state of affairs.

Consequently, Guarino et al. define a conceptualisation  $C$  to be an *intensional structure*  $(W, D, \{\mathcal{R}_j\}_{j \in \mathbb{N}})$  constituted by a non-empty set  $W$  of *possible worlds* or *states of affairs*, a non-empty set  $D$  called the *domain of discourse*, and a family of intensional relations  $\mathcal{R}_j \subseteq (2^{(D^j)})^W$ . Given a first-order signature  $\Sigma$ , a first-order language  $Sen(\Sigma)$  commits to a conceptualisation  $C$  by means of how  $\Sigma$  is interpreted into  $C$ , i.e., how predicate symbols are mapped to intensional relations. Therefore, Guarino et al. define an *ontological commitment  $K$  for  $L$*  as an intensional interpretation  $(C, I)$ , where  $I$  assigns to each relation symbol in  $\Sigma$  an intensional relation of  $C$ .

Again, analogously to what we recalled in Section 2, a conceptualisation according to Guarino et al. is a mathematical structure specified by means of a logical theory — a reference theory  $T = (\Sigma, \Gamma)$ .

Unlike Section 2, though,  $T$  is not a first-order theory, but a theory in the modal extension  $Sen^m(\Sigma)$  of the first-order language  $Sen(\Sigma)$  (i.e., the elements of  $Sen^m(\Sigma)$  are well-formed, closed first-order formulae over  $\Sigma$  that include modal operators  $\Box$  and  $\Diamond$ ); and a conceptualisation is not a first-order structure, but a structure in S5 modal logic. Guarino et al. take Kripke structures whose accessibility relation between worlds is universal (and thus can be dropped from the definition of a conceptualisation).

The map of sentences  $\alpha$  and map of structures  $\beta$  of Section 2 that restrict the set of models for first-order language  $Sen(\Sigma)$  to those first-order  $\Sigma$ -structures compatible with a conceptualisation  $C$  need now to be rethought taking into account that the ontological commitment involves two separate logical systems, namely first-order logic and modal logic. Since Guarino et al. formalise ontological commitment only in the scope of a single signature  $\Sigma$ ,  $\alpha$  amounts to the inclusion map of  $Sen(\Sigma)$  into its modal extension  $Sen^m(\Sigma)$ . The map  $\beta$  extracts from an intensional structure  $(W, D, \{\mathcal{R}_j\}_{j \in \mathbb{N}})$  the first-order structures  $(D, \{R_j^w\}_{j \in \mathbb{N}})$  for all possible worlds  $w \in W$ , where  $R_j^w = \{\rho(w) \mid \rho \in \mathcal{R}_j\}$ .

### 3.2. A General Pattern of Ontological Commitments

The general pattern is the same as before, with  $\alpha$  and  $\beta$  capturing together the commitment of language  $Sen(\Sigma)$  to a conceptualisation  $C$  — a model of reference theory  $T$  — only that the commitment is now formalised involving a first-order language and its modal extension. By contrast, the formalisation is carried out only in the scope of a single signature  $\Sigma$ . This hints at the hypothesis that formalisation of ontology commitment should not be tied to the choice of particular logical systems, maps of sentences, and maps of structures. As argued in Section 1 we actually believe in the convenience of formalising ontology commitment and semantic integration across different interpretation mechanisms: the same way Guarino et al. find a first-order characterisation of these notions insufficient, so do others not agree with a possible-worlds approach (Menzel, 2002; Santini, 2006). Still, the general pattern underlying the notion of ontological commitment that we highlighted in Sections 2 and 3 can be formalised independently of the particular choice of logical systems, maps of sentences, and maps of structures. For this we resort to institutions.

## 4. Theory of Institutions

We recall the basics of the theory of institutions that are required for the mathematical formalisation proposed in this paper. For this we shall assume some basic knowledge of category theory, particularly the notions of category, opposite category, functor, and natural transformation (see, for instance, (Goguen, 1991) for an intuitive description to the concepts of category theory in computer science, and (Pierce, 1991; Barr and Wells, 1999) for more comprehensive introductions).

Institutions originated in the late 1970s and early 1980s for studying model-theoretic properties of logics (Goguen and Burstall, 1992) and they have given semantics to powerful module systems of both imperative and declarative programming languages, multi-logic specification languages, databases, and ontologies. Most recently they have been also applied to provide abstract semantics to semantic web languages (Lucanu et al., 2006). An institution captures the essential aspects of logical systems that underlie any formal specification of a computer program: a notion of a signature system, of well-formed sentences over a signature, and for each signature, notions of a system of models and a satisfaction relation between models and sentences.

Formally, an institution is a quadruple  $\mathcal{I} = (\text{Sign}, \text{Sen}, \text{Mod}, \models)$  consisting of

- a category  $\text{Sign}$  of signatures and signature morphisms;
- a functor  $\text{Sen} : \text{Sign} \rightarrow \text{Set}$  assigning to each signature  $\Sigma$  a set of well-formed  $\Sigma$ -sentences;
- a functor  $\text{Mod} : \text{Sign}^{op} \rightarrow \text{Set}$  assigning to each signature  $\Sigma$  a set of  $\Sigma$ -structures;
- a function  $\models$  assigning to each  $\Sigma$  a binary relation  $\models_{\Sigma} \subseteq \text{Mod}(\Sigma) \times \text{Sen}(\Sigma)$  called satisfaction;

The system of first-order logic, for instance, constitutes an institution as follows:

Let  $\mathcal{I}^{\text{FOL}} = (\mathbf{Sign}^{\text{FOL}}, \text{Sen}^{\text{FOL}}, \text{Mod}^{\text{FOL}}, \models^{\text{FOL}})$  such that  $\mathbf{Sign}^{\text{FOL}}$  is the category of first-order signatures and their morphisms, and for a signature  $\Sigma$  in  $\mathbf{Sign}^{\text{FOL}}$ ,

- $\text{Sen}^{\text{FOL}}(\Sigma)$  is the set of all well-formed, closed first-order formulae over  $\Sigma$ ;
- $\text{Mod}^{\text{FOL}}(\Sigma)$  is the set of all structures  $\mathfrak{S} = (D, \{\mathcal{R}_j\}_{j \in \mathbb{N}})$  with nonempty domain of discourse  $D$  and family of sets of relations  $\mathcal{R}_j = \{r^{\mathfrak{S}} \subseteq D^j\}$  where  $r^{\mathfrak{S}}$  stands for the relation assigned by  $\mathfrak{S}$  to the  $j$ -ary relation symbol  $r \in \Sigma$ ; and
- $\models_{\Sigma}^{\text{FOL}}$  is first-order satisfaction of  $\Sigma$ -structures for  $\Sigma$ -sentences.

Analogously the system of S5 modal logic also constitutes an institution as follows:

Let  $\mathcal{I}^{\text{S5}} = (\mathbf{Sign}^{\text{S5}}, \text{Sen}^{\text{S5}}, \text{Mod}^{\text{S5}}, \models^{\text{S5}})$  such that  $\mathbf{Sign}^{\text{S5}}$  is the same category  $\mathbf{Sign}^{\text{FOL}}$  of first-order signatures and their morphisms, and for a signature  $\Sigma$  in  $\mathbf{Sign}^{\text{S5}}$ ,

- $\text{Sen}^{\text{S5}}(\Sigma)$  is the set of all well-formed, closed first-order formulae over  $\Sigma$  including the modality operators  $\Box$  and  $\Diamond$ ;
- $\text{Mod}^{\text{S5}}(\Sigma)$  is the set of all pairs  $(w, \mathfrak{S})$ , where  $\mathfrak{S} = (W, D, \{\mathcal{R}_j\}_{j \in \mathbb{N}})$  is an intensional structure with nonempty set of possible worlds  $W$ , nonempty domain of discourse  $D$ , and family of sets of intensional relations  $\mathcal{R}_j = \{r^{\mathfrak{S}} \subseteq (2^{D^j})^W\}$  where  $r^{\mathfrak{S}}$  stands for the intensional relation assigned by  $\mathfrak{S}$  to the  $j$ -ary relation symbol  $r \in \Sigma$ , and  $w \in W$ ; and
- $\models_{\Sigma}^{\text{S5}}$  is first-order satisfaction of  $\Sigma$ -structures for  $\Sigma$ -sentences in possible worlds.

It is easy to check that  $\mathcal{I}^{\text{FOL}}$  and  $\mathcal{I}^{\text{S5}}$  are both well-defined and satisfy the fundamental property of institutions.

Fig. 4. The institutions of first-order and S5 modal logic.

satisfying the following fundamental property: For all  $\sigma : \Sigma \rightarrow \Sigma'$ ,  $\varphi \in \text{Sen}(\Sigma)$  and  $M' \in \text{Mod}(\Sigma')$ ,

$$\text{Mod}(\sigma)(M') \models_{\Sigma} \varphi \quad \text{iff} \quad M' \models_{\Sigma'} \text{Sen}(\sigma)(\varphi) .$$

Notice that  $\text{Sen}(\sigma) : \text{Sen}(\Sigma) \rightarrow \text{Sen}(\Sigma')$  is the function translating  $\Sigma$ -sentences to  $\Sigma'$ -sentences, while  $\text{Mod}(\sigma) : \text{Mod}(\Sigma') \rightarrow \text{Mod}(\Sigma)$  is the function translating  $\Sigma'$ -structures to  $\Sigma$ -structures. Figure 4 shows institutions for first-order and S5 modal logic.

A theory is a pair  $(\Sigma, \Gamma)$  where  $\Sigma$  is an object in  $\mathbf{Sign}$  and  $\Gamma \subseteq \text{Sen}(\Sigma)$ . Given a theory  $(\Sigma, \Gamma)$ , let  $\text{Mod}(\Sigma, \Gamma)$  be the subset of  $\text{Mod}(\Sigma)$  determined by those models  $M \in \text{Mod}(\Sigma)$  for which  $M \models_{\Sigma} \psi$ , for all  $\psi \in \Gamma$ . We write  $\Gamma \models_{\Sigma} \varphi$  if, for all  $M \in \text{Mod}(\Sigma, \Gamma)$ ,  $M \models_{\Sigma} \varphi$ . A theory morphism  $\sigma : (\Sigma, \Gamma) \rightarrow (\Sigma', \Gamma')$  is a signature morphism  $\sigma : \Sigma \rightarrow \Sigma'$  such that, for all  $\varphi \in \text{Sen}(\Sigma)$ ,  $\Gamma \models_{\Sigma} \varphi$  implies  $\Gamma' \models_{\Sigma'} \text{Sen}(\sigma)(\varphi)$ . Let  $\mathbf{Th}$  be the category of theories and theory morphisms.

We extend  $\models$  to a function assigning to a theory  $T = (\Sigma, \Gamma)$  a binary relation  $\models_T \subseteq \models_{\Sigma}$ , such that  $M \models_T \varphi$  if and only if,  $M \in \text{Mod}(T)$  and  $M \models_{\Sigma} \varphi$ . We also extend the functor  $\text{Sen} : \mathbf{Sign} \rightarrow \mathbf{Set}$  to a functor  $\text{Sen} : \mathbf{Th} \rightarrow \mathbf{Set}$  by composing it with the forgetful functor  $\text{sign} : \mathbf{Th} \rightarrow \mathbf{Sign}$  that forgets axioms.

Given institutions  $\mathcal{I} = (\mathbf{Sign}, \text{Sen}, \text{Mod}, \models)$  and  $\mathcal{I}' = (\mathbf{Sign}', \text{Sen}', \text{Mod}', \models')$ , a *map of institutions*<sup>3</sup>  $F : \mathcal{I} \rightarrow \mathcal{I}'$  is a triple  $F = (\Phi, \alpha, \beta)$  consisting of

- a functor  $\Phi : \mathbf{Sign} \rightarrow \mathbf{Sign}'$ ;
- a natural transformation  $\alpha : \text{Sen} \Rightarrow \text{Sen}' \circ \Phi$ ;
- a natural transformation  $\beta : \text{Mod}' \circ \Phi^{op} \Rightarrow \text{Mod}$

satisfying the following fundamental property: For all  $\Sigma \in \mathbf{Sign}$ ,  $\varphi \in \text{Sen}(\Sigma)$  and  $M' \in \text{Mod}'(\Phi(\Sigma))$ ,

$$\beta_{\Sigma}(M') \models_{\Sigma} \varphi \quad \text{iff} \quad M' \models_{\Phi(\Sigma)} \alpha_{\Sigma}(\varphi) .$$

<sup>3</sup>Our definition of a map of institutions is close to Meseguer's one — it actually is equivalent to that of a *simple map of institutions* (Meseguer, 1989). It corresponds also to what Goguen and Roşu call a *simple theoroidal comorphism of institutions* (Goguen and Roşu, 2002).

---

Let  $F : \mathcal{I}^{\text{FOL}} \rightarrow \mathcal{I}^{\text{S5}}$  be the map of institutions  $F = (\Phi, \alpha, \beta)$ , such that for a signature  $\Sigma$  in  $\text{Sign}^{\text{FOL}}$ ,

- $\Phi(\Sigma) = (\Sigma, \emptyset)$ ;
- $\alpha_\Sigma$  is the inclusion map of  $\text{Sen}^{\text{FOL}}(\Sigma)$  into  $\text{Sen}^{\text{S5}}(\Sigma)$ ; and
- given a pair  $(w, \mathfrak{S})$  in  $\text{Mod}^{\text{S5}}(\Phi(\Sigma))$  with  $\mathfrak{S} = (W, D, \{\mathcal{R}_j\}_{j \in \mathbb{N}})$ ,  $\beta_\Sigma(w, \mathfrak{S}) = (D, \{R_j\}_{j \in \mathbb{N}})$ , where  $R_j = \{r^{\mathfrak{S}}(w) \mid r \text{ is an } j\text{-ary relation symbol in } \Sigma\}$ .

It is easy to check that this map is well defined and indeed satisfies the fundamental property of a map of institutions.

Fig. 5. A map from the institution of first-order logic to the institution of S5 modal logic.

---

Figure 5 shows a map from the institution of first-order logic to the institution of S5 modal logic.<sup>4</sup>

## 5. Institutionalising Ontological Commitment

We now unfold a mathematical formalisation of language, ontology, and ontological commitment in terms of institutions that accounts for the general syntax- and interpretation-independent pattern put forward in Sections 2 and 3. For all the subsequent definitions, let  $\mathcal{I} = (\text{Sign}, \text{Sen}, \text{Mod}, \models)$  and  $\mathcal{I}' = (\text{Sign}', \text{Sen}', \text{Mod}', \models')$  be two (not necessarily distinct) institutions.

### 5.1. Ontology for a Language

Under a language  $L$  one usually understands the set of all sentences that can be formed over a particular vocabulary. The way sentences are formed, though, should not be conditioned by the choice of a particular vocabulary, but by the institution, for instance when we are in the context of a first-order language. The same happens with the interpretation of the language. Although the class of structures endowing a language with a semantics is determined by a particular choice of vocabulary, the way structures are formed should not be conditioned by this choice. This is captured by both  $\text{Sen}$  and  $\text{Mod}$  in an institution  $\mathcal{I}$  being functors:

**Definition 3 (Language)** *A language  $L$  in  $\mathcal{I}$  is a set of sentences  $L = \text{Sen}(\Sigma)$  over a signature  $\Sigma$  in  $\text{Sign}$ . The signature  $\Sigma$  is often also called the vocabulary of  $L$ . By virtue of  $L$  being a language in an institution  $\mathcal{I}$  we can interpret it, which yields  $\text{Mod}(\Sigma)$ , the set of (unconstrained) models of language  $L$  whose vocabulary is  $\Sigma$ .*

We have seen in Section 2 that under an ontology one usually understands a specification of a conceptualisation for interpreting a particular language vocabulary. But in Section 3 we have also seen that such specification is not necessarily written in the context of the same institution as the one of the language the ontology is for. Consequently, we define an ontology for a language on top of an underlying map of institutions (which, in case of identical institutions, will be just an endomap):

**Definition 4 (Ontology)** *An ontology  $O$  for a language  $L = \text{Sen}(\Sigma)$  in  $\mathcal{I}$  is a theory  $O = (\Sigma', \Gamma')$  in  $\mathcal{I}'$  whenever there exists a map of institutions  $F : \mathcal{I} \rightarrow \mathcal{I}'$  with  $F = (\Phi, \alpha, \beta)$  such that  $\Phi(\Sigma) = O$ . By virtue of  $O$  being a theory in an institution  $\mathcal{I}'$  we can interpret it, which yields  $\text{Mod}'(O)$ , the class of models of  $O$ . This class constitutes the conceptualisation to which  $L$  commits by virtue of using the ontology  $O$ .*

At first sight this might seem a non-standard definition of ontology, although an ontology is still seen as a logical theory. With this definition we want to stress that we do not view an ontology as something that stands in isolation, but that provides interpretation *for* some language, and that this interpretation is soundly founded on top of a map of institutions, as we shall see next.

---

<sup>4</sup>Note that this is not the standard translation of modal logic into first-order logic, since we define a map in the other direction, from first-order logic into modal logic.

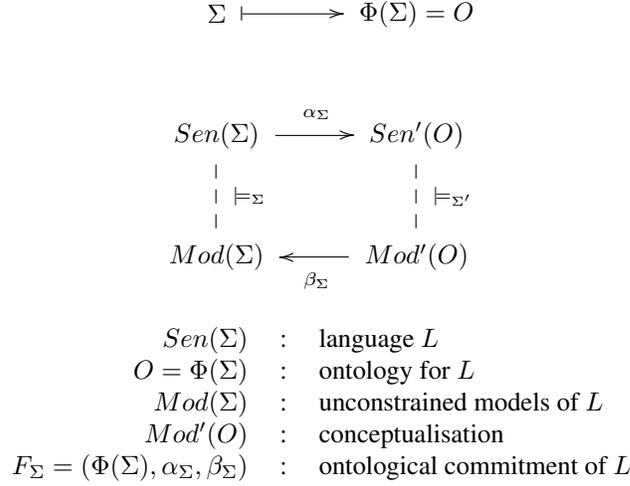


Fig. 6. Language, ontology, and ontological commitment.

## 5.2. Ontological Commitment of a Language

An ontological commitment of a language establishes the link between the language and an ontology for that language at both the syntactic and the semantic level. Thus, it interprets the language vocabulary in terms of the ontology, and it also extracts admissible models for that language from the conceptualisation specified by the ontology. The map of institutions underlying the fact that the ontology is *for* a language, provides the technical apparatus to define this commitment:

**Definition 5 (Ontological Commitment)** *A language  $L$  in  $\mathcal{I}$  with vocabulary  $\Sigma$  commits to a conceptualisation  $Mod'(O)$  specified by means of ontology  $O$  in  $\mathcal{I}'$  whenever  $O$  is an ontology for  $L$  by virtue of a map of institutions  $F : \mathcal{I} \rightarrow \mathcal{I}'$  with  $F = (\Phi, \alpha, \beta)$ , according to Definition 4. In this case the function  $\alpha_\Sigma : Sen(\Sigma) \rightarrow Sen'(O)$  defines the vocabulary of  $L$  in terms of the vocabulary of  $O$  and  $\beta_\Sigma$  determines the ontologically committed models of  $L$ , which are linked to the conceptualisation that the ontology  $O$  specifies. Thus, for a particular language  $L$  with vocabulary  $\Sigma$ , the ontological commitment is captured by means of the triple  $F_\Sigma = (\Phi(\Sigma), \alpha_\Sigma, \beta_\Sigma)$  (recall that  $O = \Phi(\Sigma)$ ).*

Figure 6 summarises the definitions above. It resembles the left (and also the right) part of Figure 3, but, while previously all notions were stated in the context of the institution of first-order logic, now we are lifting them into the syntax- and interpretation-independent framework provided by institutions. We view the interpretation of a language vocabulary *and* the extraction of models from the conceptualisation now as a whole, as the ontological commitment.

## 5.3. Semantic Integration

A syntax- and interpretation-independent characterisation of language, ontology and ontological commitment brings forth a generalisation of the notions of ‘theory interpretation,’ and ‘semantic integration’ introduced in Section 2 for the first-order case:

**Definition 6 (Theory Interpretation)** *Let  $\mathcal{I} = (\mathbf{Sign}, Sen, Mod, \models)$  and  $\mathcal{I}' = (\mathbf{Sign}', Sen', Mod', \models')$  be two (not necessarily distinct) institutions, and let  $F : \mathcal{I} \rightarrow \mathcal{I}'$  be a map of institutions with  $F = (\Phi, \alpha, \beta)$ . Given a signature  $\Sigma$  in  $\mathbf{Sign}$ , the function  $\alpha_\Sigma : Sen(\Sigma) \rightarrow Sen'(\Phi(\Sigma))$  is also a theory interpretation  $\alpha_\Sigma : T \rightarrow T'$  between theories  $T = (\Sigma, \Gamma)$  in  $\mathbf{Th}$  and  $T' = (\Sigma', \Gamma')$  in  $\mathbf{Th}'$  (where  $\Sigma'$  is the signature of  $\Phi(\Sigma)$ ) if, and only if, for all  $\varphi \in Sen(\Sigma)$ ,  $\Gamma \models_\Sigma \varphi$  implies  $\Gamma' \models_{\Phi(\Sigma)} \alpha_\Sigma(\varphi)$ . It is a faithful theory interpretation if, in addition,  $\Gamma' \models_{\Phi(\Sigma)} \alpha_\Sigma(\varphi)$  implies  $\Gamma \models_\Sigma \varphi$ .*

**Definition 7 (Semantic Integration)** Let  $\mathcal{T}^1 = (\mathbf{Sign}^1, \mathbf{Sen}^1, \mathbf{Mod}^1, \models^1)$ ,  $\mathcal{T}^2 = (\mathbf{Sign}^2, \mathbf{Sen}^2, \mathbf{Mod}^2, \models^2)$  and  $\mathcal{I} = (\mathbf{Sign}, \mathbf{Sen}, \mathbf{Mod}, \models)$  be three (not necessarily distinct) institutions. We say that two theories  $T_1 = (\Sigma_1, \Gamma_1)$  and  $T_2 = (\Sigma_2, \Gamma_2)$  in  $\mathbf{Th}^1$  and  $\mathbf{Th}^2$  are semantically integrated with respect to  $T$  in  $\mathbf{Th}$ , if

- $T$  is an ontology for both languages  $\mathbf{Sen}^1(\Sigma_1)$  and  $\mathbf{Sen}^2(\Sigma_2)$ , with ontological commitments  $F_{\Sigma_i}^i = (T, \alpha_{\Sigma_i}^i, \beta_{\Sigma_i}^i)_{i=1,2}$ ;
- the functions  $\alpha_{\Sigma_1}^1$  and  $\alpha_{\Sigma_2}^2$  are also theory interpretations  $\alpha_{\Sigma_1}^1 : T_1 \rightarrow T$  and  $\alpha_{\Sigma_2}^2 : T_2 \rightarrow T$ ;
- $\mathbf{Mod}(T) \neq \emptyset$ .

We call  $\mathfrak{J} = \{\alpha_{\Sigma_i}^i : T_i \rightarrow T; \beta_{\Sigma_i}^i : \mathbf{Mod}(T) \rightarrow \mathbf{Mod}^i(T_i)\}_{i=1,2}$  the semantic integration of  $T_1$  and  $T_2$  with respect to  $T$  on the basis of ontological commitments  $F_{\Sigma_i}^i = (T, \alpha_{\Sigma_i}^i, \beta_{\Sigma_i}^i)_{i=1,2}$ .

As before, we shall call  $\mathfrak{J}$  a *faithful* semantic integration if its theory interpretations are faithful. Note that, in the general case, theory interpretation and semantic integration occur relative to one or several ontological commitments, which in turn are based on one or several maps of institutions. Analogously, the notion of ‘ontology-based consequence’ can also be lifted into the institutional framework:

**Definition 8 (Ontology-Based Consequence)** Let  $\mathfrak{J} = \{\alpha_{\Sigma_i}^i : T_i \rightarrow T; \beta_{\Sigma_i}^i : \mathbf{Mod}(T) \rightarrow \mathbf{Mod}^i(T_i)\}_{i=1,2}$  be a semantic integration of  $T_1 = (\Sigma_1, \Gamma_1)$  and  $T_2 = (\Sigma_2, \Gamma_2)$  with respect to  $T$ , and let  $\varphi \in \mathbf{Sen}^1(\Sigma_1)$  and  $\Gamma \subseteq \mathbf{Sen}^2(\Sigma_2)$ . We say that  $\varphi$  is an ontology-based consequence of  $\Gamma$ , written  $\Gamma \models_{\mathfrak{J}} \varphi$ , if  $\alpha_{\Sigma_2}^2(\Gamma) \models_T \alpha_{\Sigma_1}^1(\varphi)$ .

## 6. Eliciting Ontological Commitments

We now take four increasingly complex notions of semantic integration and ontological commitment that use differing understandings of semantics and use our institutionalised framework to elicit and compare their underlying ontological commitments. First, we take Stumme and Maedche’s ontology merging method, FCA-Merge (Stumme and Maedche, 2001), where the semantics of a concept symbol is captured through the instances classified by that symbol. Next, we look at Bench-Capon, Malcolm and Shave’s formalisation of relations between ontologies based on homomorphisms of order-sorted algebras (Bench-Capon and Malcolm, 1999; Bench-Capon et al., 2003). Further, we analyse Grüninger and Kopena’s interlingua approach using the Process Specification Language (PSL), which is based on first-order structure invariants (Grüninger and Kopena, 2005). Finally, we revisit Guarino, Carrara and Giaretta’s formalisation of ontological commitment discussed in Section 3, which uses possible-world semantics (Guarino et al., 1994; Guarino, 1998a). Table 1 summarises the analysis carried out below.

	<i>institution(s)</i>	<i>ontology</i>	<i>semantics</i>	<i>theory interpretations</i>
(Stumme and Maedche, 2001)	formal contexts	partial order	instance-based	order-preserving maps
(Bench-Capon and Malcolm, 1999), (Bench-Capon et al., 2003)	order-sorted, conditional equational logic	conditional equational theory	order-sorted algebras	theory morphisms
(Grüninger and Kopena, 2005), (Ciocoiu and Nau, 2000)	first-order logic	first-order theory	first-order structures	endomaps of institutions
(Guarino et al., 1994), (Guarino, 1998a)	first-order logic and S5 modal logic	S5 modal theory	possible worlds	map of institutions

Table 1  
Increasingly complex semantic integration formalisms

### 6.1. Instance-Based Ontology Merging

FCA-Merge is a method for merging two concept hierarchies  $C_1$  and  $C_2$  to compute a new hierarchy using techniques from formal concept analysis (Ganter and Wille, 1999). Merging is done relative to a fixed set of documents  $D$  that are used as instances to be classified with respect to concepts. This classification is done automatically by means of a linguistic analysis of the documents, classifying a particular document to those concept symbols that occur in the document. This yields two classification relations  $K_1 \subseteq D \times C_1$  and  $K_2 \subseteq D \times C_2$  called *formal contexts*. These two formal contexts are joined, yielding a new formal context  $K = (K_1 \uplus K_2) \subseteq D \times (C_1 \uplus C_2)$ . (The symbols ‘ $\uplus$ ’ stands for the disjoint union of sets.) In turn this new formal context is used to generate a so called *concept lattice*. This lattice can serve as a guide for the knowledge engineer to manually construct the new concept hierarchy out of the merged concepts  $C_1 \uplus C_2$ .

The integration is carried out in the context of a single institution  $\mathcal{I}^{\text{FCA}} = (\text{Sign}^{\text{FCA}}, \text{Sen}^{\text{FCA}}, \text{Mod}^{\text{FCA}}, \models^{\text{FCA}})$ , where  $\text{Sign}^{\text{FCA}}$  is the category of partially ordered sets and order-preserving maps representing concept hierarchies;  $\text{Sen}^{\text{FCA}}$  is the identity functor, i.e., for a concept hierarchy  $C$ ,  $\text{Sen}^{\text{FCA}}(C) = C$  (sentences are the concept symbols themselves);  $\text{Mod}^{\text{FCA}}$  assigns to a concept hierarchy  $C$  a set  $\text{Mod}^{\text{FCA}}(C)$  of instances (a set of documents); and  $\models^{\text{FCA}}$  is the classification relation of instances (documents) in  $\text{Mod}^{\text{FCA}}(C)$  to concept symbols in  $C$  that respects the concept hierarchy.

FCA-Merge is based on the assumption that two concept hierarchies  $C_1$  and  $C_2$  share the same set of instances  $D$ , i.e.,  $\text{Mod}^{\text{FCA}}(C_1) = \text{Mod}^{\text{FCA}}(C_2) = D$ . Taking as local theories  $T_1 = (C_1, \emptyset)$  and  $T_2 = (C_2, \emptyset)$  and as reference theory  $T = (C_1 \uplus C_2, \emptyset)$ , the order-embeddings  $\sigma_1 : C_1 \rightarrow C_1 \uplus C_2$  and  $\sigma_2 : C_2 \rightarrow C_1 \uplus C_2$  of the concept hierarchies into the disjoint union are obviously also theory morphisms (and hence theory interpretations)  $\sigma_1 : T_1 \rightarrow T$  and  $\sigma_2 : T_2 \rightarrow T$ . Consequently, they constitute a semantic integration on the basis of ontological commitments  $F_{C_i} = ((C_1 \uplus C_2, \emptyset), \sigma_i, id_D)_{i=1,2}$ , where  $id_D$  is the identity map on the set of instances  $D$ .

### 6.2. Generic Ontology and Schema Interoperability

Bench-Capon, Malcolm and Shave were probably the first to give an explicit formalisation of ontologies and their relations between them using institutions. They extend the concept of abstract data type to that of ontology, thus building upon the universal algebra tradition. As with FCA-Merge, integration is carried out in the context of a single institution, in particular it is the institution  $\mathcal{I}^{\text{CEQL}} = (\text{Sign}^{\text{CEQL}}, \text{Sen}^{\text{CEQL}}, \text{Mod}^{\text{CEQL}}, \models^{\text{CEQL}})$  of order-sorted conditional equational logic. Signatures in  $\text{Sign}^{\text{CEQL}}$  are order-sorted signatures of unary function symbols. For a signature  $\Sigma$ ,  $\text{Sen}^{\text{CEQL}}(\Sigma)$  is the set of conditional equations over  $\Sigma$ ,  $\text{Mod}^{\text{CEQL}}(\Sigma)$  is the set of order-sorted  $\Sigma$ -algebras, and  $\models^{\text{CEQL}}$  is the usual satisfaction relation between algebras and conditional equations.

Ontologies are formalised as theories in  $\text{Th}^{\text{CEQL}}$ , and their interoperability is characterised by means of a *relation between two theories*  $T_1 = (\Sigma_1, \Gamma_1)$  and  $T_2 = (\Sigma_2, \Gamma_2)$  in  $\text{Th}^{\text{CEQL}}$ . These are formalised with two theory morphisms  $\chi_1 : T_0 \rightarrow T_1$  and  $\chi_2 : T_0 \rightarrow T_2$  sharing their source  $T_0 = (\Sigma_0, \Gamma_0)$ . Such a relation between two theories can be seen as an alternative formalisation of their semantic integration: for institution  $\mathcal{I}^{\text{CEQL}}$  the category  $\text{Th}^{\text{CEQL}}$  is *cocomplete*, which means that a relation  $\chi_i : T_0 \rightarrow T_i$  ( $i = 1, 2$ ) between two theories always has a pushout<sup>5</sup>  $\sigma_i : T_i \rightarrow T$ . This pushout, obviously, constitutes a semantic integration on the basis of ontological commitments  $F_{\Sigma_i} = (T, \text{Sen}^{\text{CEQL}}(\sigma_i), \text{Mod}^{\text{CEQL}}(\sigma_i))_{i=1,2}$ .

By formalising ontology interoperability with institutions, Bench-Capon et al. hint at the fact that one is not restricted to remain confined in the context of order-sorted conditional equational logic. Kent, for instance, has adopted institutions to represent and organise ontological structures within his *Information Flow Framework* (IFF) (Kent, 2000) in order to be independent of the particular logic used in ontologies. Furthermore, he has also proposed a formal characterisation of semantic integration in IFF in an institutionalised fashion (Kent, 2005). Still, in IFF semantic integration is described, like in Bench-Capon et

<sup>5</sup>A pushout is a generalised disjoint union where a common part is identified (see also (Mac Lane, 1998)).

al.’s approach, in the context of a single institution — albeit still in a logic-independent fashion as no particular institution is favoured. Analogously, Alagić and Bernstein also use institutions as a foundation of their model theory for generic management of database schemata that is independent of a particular data model (Alagić and Bernstein, 2002). Again, their framework is based on a single institution, which they call *schema transformation framework*.

### 6.3. Translation through an Interlingua

Grüniger and Kopena describe an interlingua approach to semantic integration based on first-order model-theoretic semantics. They exemplify it with the Process Specification Language (PSL), an ontology that has been designed using *invariants*, i.e., properties of models that are preserved under isomorphism. Local ontologies are mapped to PSL by running a so called “twenty-questions tool” that attempts to identify those invariants that are or are not preserved by local models. Local class symbols are then defined using PSL terminology according to the answers to these “twenty questions.”

Like the previous two cases of Sections 6.1 and 6.2, semantic integration involves a single institution. Here it is the institution  $\mathcal{I}^{\text{FOL}}$  of first-order logic. The PSL ontology is a theory  $T_{PSL} = (\Sigma_{PSL}, \Gamma_{PSL})$  in  $\text{Th}^{\text{FOL}}$ . Unlike the scenarios before, however, the integration is not based on signature or theory morphisms in  $\text{Sign}^{\text{FOL}}$  or  $\text{Th}^{\text{FOL}}$ , respectively. A particular atomic sentence with a predicate in the signature  $\Sigma$  of local ontology  $T$  may be mapped to a non-atomic first-order formula in  $\text{Sen}^{\text{FOL}}(\Sigma_{PSL})$ . The “twenty-question tool” described in (Grüniger and Kopena, 2005) assists a knowledge engineer in defining the map  $\alpha_{\Sigma} : \text{Sen}^{\text{FOL}}(\Sigma) \rightarrow \text{Sen}^{\text{FOL}}(\Sigma_{PSL})$ . This map is then the component on  $\Sigma$  of the natural transformation  $\alpha$  of an endomap  $F : \mathcal{I}^{\text{FOL}} \rightarrow \mathcal{I}^{\text{FOL}}$  between the institution of first-order logic and itself. Consequently, Grüniger and Kopena’s “twenty questions” for local terminology  $\Sigma$  determine an ontological commitment  $F_{\Sigma} = (T_{PSL}, \alpha_{\Sigma}, \beta_{\Sigma})$ , where  $\beta_{\Sigma}$  is the map of first-order structures determined by  $\alpha_{\Sigma}$  as described in Section 2.

Ciociu and Nau’s formalisation of *ontology-based partial translation* in (Ciociu and Nau, 2000) is also an example of an ontology commitment with a first-order logic endomap as underlying map of institutions.

### 6.4. Ontological Commitment, Revisited

The institutionalised notion of ‘ontological commitment’ of Definition 5 relates to Guarino, Carrara and Giaretta’s formalisation of the same notion discussed in Section 3 through the fact that the latter arises as a special case of the former by fixing a map of institutions from the institution of first order logic to the institution of S5 modal logic. We use for this the *internal language*  $\text{Lang}(C)$  for an intensional structure  $C$ , described next, and model the intensional interpretation  $K = (C, I)$  that formalises Guarino et al.’s ontological commitment for a language  $L$  with a map from  $L$  to  $\text{Lang}(C)$  characterised by the natural transformation  $\alpha$  of the map of institutions from first-order logic to S5 modal logic. The internal language  $\text{Lang}(C)$  for an intensional structure  $C = (W, D, \{\mathcal{R}_j\}_{j \in \mathbb{N}})$  is obtained by defining a signature  $\Sigma_C$  such that  $C$  is a  $\Sigma_C$ -structure. The obvious way to do this is by taking the elements of  $\mathcal{R}_i$  themselves as relation symbols of  $\Sigma_C$ ;  $\text{Lang}(C)$  is then  $\text{Sen}^{\text{S5}}(\Sigma_C)$ . The map  $I$  assigning to each relation symbol in  $\Sigma$  an intensional relation of  $C$  can then also be seen as a signature morphism from  $\Sigma$  to  $\Sigma_C$ .

Let  $\mathcal{I}^{\text{FOL}}$  and  $\mathcal{I}^{\text{S5}}$  be the institutions of Figure 4 and let us take now the map of institutions  $(\Phi, \alpha, \beta) : \mathcal{I}^{\text{FOL}} \rightarrow \mathcal{I}^{\text{S5}}$ , such that for signature  $\Sigma$  in  $\text{Sign}^{\text{FOL}}$

- $\Phi(\Sigma) = (\Sigma_C, \emptyset)$ ;
- $\alpha_{\Sigma} = \text{Sen}^{\text{FOL}}(I)$  is the map from  $\text{Sen}^{\text{FOL}}(\Sigma)$  into  $\text{Lang}(C)$ ; and
- given a pair  $(w, \mathfrak{S})$  in  $\text{Mod}^{\text{S5}}(\Phi(\Sigma))$  with  $\mathfrak{S} = (W, D, \{\mathcal{R}_j\}_{j \in \mathbb{N}})$ ,  $\beta_{\Sigma}(w, \mathfrak{S}) = (D, \{R_j\}_{j \in \mathbb{N}})$ , where  $R_j = \{r^{\mathfrak{S}}(w) \mid r \text{ is an } j\text{-ary relation symbol in } \Sigma\}$ .

$F_\Sigma = (\Phi(\Sigma), \alpha_\Sigma, \beta_\Sigma)$  is an ontological commitment for  $L = \text{Sen}^{\text{FOL}}(\Sigma)$  according to Definition 5, and we get Guarino et al.’s ontological commitment by taking as intensional interpretation  $K = (C, \alpha_\Sigma)$ ,<sup>6</sup> where  $C \in \text{Mod}^{\text{S5}}(\Phi(\Sigma))$ . If  $\Phi(\Sigma)$  was categorical, i.e., if all models in  $\text{Mod}^{\text{S5}}(\Phi(\Sigma))$  were isomorphic, then  $\Phi(\Sigma)$  would uniquely specify the conceptualisation  $C$  (up to isomorphism). Generally, this is not the case, and hence  $\Phi(\Sigma)$  only approximately specifies  $C$ .  $\Phi(\Sigma)$  is the ontology for  $L = \text{Sen}^{\text{FOL}}(\Sigma)$ .

## 7. Conclusion

The search for a suitable mathematical framework on which to build a genuine engineering discipline of semantic integration has brought various researchers to look at category theory for the necessary foundations (Jannink et al., 1998; Kent, 2000; Hitzler et al., 2005; Zimmermann et al., 2006). Category theory has provided deep insights in the fields of mathematical logic and computer science, and it has often been used as a guide for finding good definitions and research directions (Goguen, 1991). Its abstractness, however, has often been seen as not directly useful to the knowledge engineer or practitioner that needs to address practical semantic heterogeneity problems (Menzel, 2005). But, the variety of logical systems used in knowledge engineering, and their associated interpretation mechanisms reflecting different understandings of semantics, calls for a framework that is general enough to cope with heterogeneity both in terminology *and* in interpretation mechanism, but that is also concrete enough for providing insights into the actual deployment of semantic integration technology.

In this paper we have explored the suitability of institutions as an adequate category-theoretic tool for providing both general and useful definitions to address the semantic heterogeneity problem. In Definition 5 we used institutions to define the ontological commitment of a language with respect to a particular choice of vocabulary  $\Sigma$ . But the institutional framework allows us to consider the general act of an ontological commitment not to be conditioned by the particular choice of vocabulary. Hence, a particular ontological commitment of a language arises as an element of a family of commitments. By virtue of defining the ontological commitment of a language on top of a map of institutions  $F = (\Phi, \alpha, \beta)$ , the general act of committing is captured through the fact that  $\Phi$  is a functor and that both  $\alpha$  and  $\beta$  are natural transformations.

The additional gain we get from deriving a general notion of ontological commitment from that of a map of institutions is that we can now describe the relationship between ontological commitments for each particular language vocabulary from the relationship between vocabularies as captured in signature morphisms in the category  $\text{Sign}$ . This, for instance, may be useful for determining the ontological commitment that arises when one uses, in a modular fashion, various ontologies for separate fragments of the vocabulary of a same language. For first-order ontologies, Lüttich, Mossakowski, and Krieg-Brückner, exploit such algebraic approach with specifications in CASL.<sup>7</sup> This enables them to apply the tool set HETS (Mossakowski et al., 2007) to carry out syntax and type analysis using the connection to various provers based on the combination and detection of various different logics and sublogics (Lüttich et al., 2004; Lüttich and Mossakowski, 2004).

A limitation of this move, however, is that an ontological commitment is not just adding information and selecting models but adding a “certain kind of information” and selecting “certain models.” When choosing a particular formalisation of ontological commitment such as the one carried out by Guarino, Carrara and Giaretta using an intensional structure, this “certain kind of information” is given an intensional character, a particularity that is not retained in our abstract institutional framework. So probably our framework as it stands now overgeneralises certain issues that might be of central importance from the ontological point of view, and also from the more pragmatic engineering perspective. Consequently, using the *full-blown* theory of institutions is probably not yet at the right level of abstraction to make it directly suitable for knowledge engineers and practitioners. We still need a theoretical framework that fixes the

<sup>6</sup>Rigorously speaking  $\alpha_\Sigma$  is not a map of signature symbols to intensional relations as required; it can be seen as such, however, if we take its restriction to atomic formulae.

<sup>7</sup>The Common Algebraic Specification Language designed by the Common Framework Initiative (CoFI) (Mosses, 2004).

fundamental ideas of the institutional framework while hiding the category-theoretic machinery. Category theory, and thus institutions, are unfamiliar even to most mathematicians and logicians. No one developing semantic integration methods and tools should therefore ever be faced with institutions directly, but they should be able to work with a rigorous theory of semantic integration that was built upon it. We only hope that the framework described in this article serves a step further in building such a foundation.

### **Acknowledgements**

We are very grateful to the reviewers of this article. Their detailed comments and constructive suggestions have been very valuable to improve and build upon previous versions.

This work is supported under the UPIC project, sponsored by Spain's Ministry of Education and Science under grant number TIN2004-07461-C02-02; under the OpenKnowledge Specific Targeted Research Project (STREP), sponsored by the European Commission under contract number FP6-027253; and under the Advanced Knowledge Technologies (AKT) Interdisciplinary Research Collaboration (IRC), sponsored by the UK Engineering and Physical Sciences Research Council under grant number GR/N15764/01. M. Schorlemmer is also supported by a *Ramón y Cajal* Research Fellowship from Spain's Ministry of Education and Science, which is partially funded by the European Social Fund.

## References

- Alagić, S. and Bernstein, P. (2002). A model theory for generic schema management. In Ghelli, G. and Grahne, G., editors, *Database Programming Languages. 8th International Workshop, DBPL 2001, Frascati, Italy, September 8-10, 2001. Revised Papers*, volume 2397 of *Lecture Notes in Computer Science*, pages 228–246. Springer.
- Atencia, M. and Schorlemmer, M. (2007). A formal model for situated semantic alignment. In Durfee, E. H. and Yokoo, M., editors, *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems. May 14–18, 2007. Honolulu, Hawai'i*, pages 1270–1277. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).
- Barr, M. and Wells, C. (1999). *Category Theory for Computing Science*. Centre de Recherches Mathématiques, Université de Montreal, third edition.
- Bench-Capon, T. and Malcolm, G. (1999). Formalising ontologies and their relations. In Bench-Capon, T., Soda, G., and Tjoa, A. M., editors, *Database and Expert Systems Applications. 10th International Conference, DEXA '99, Florence, Italy, August 30 - September 3, 1999. Proceedings*, volume 1677 of *Lecture Notes in Computer Science*, pages 250–259. Springer.
- Bench-Capon, T., Malcolm, G., and Shave, M. (2003). Semantics for interoperability: relating ontologies and schemata. In Marík, V., Retschitzegger, W., and Stepánková, O., editors, *Database and Expert Systems Applications. 14th International Conference, DEXA 2003, Prague, Czech Republic, September 1-5, 2003. Proceedings*, volume 2736 of *Lecture Notes in Computer Science*, pages 703–712. Springer.
- Benjamins, R., Euzenat, J., Noy, N., Shvaiko, P., Stuckenschmidt, H., and Uschold, M., editors (2006). *International Workshop on Ontology Matching. Co-located with the 5th International Semantic Web Conference ISWC-2006, November 5, 2006, Athens, Georgia, USA*.
- Borst, P., Akkermans, H., and Top, J. (1997). Engineering ontologies. *International Journal of Human-Computer Studies*, 46(2–3):365–406.
- Calvanese, D. and De Giacomo, G. (2005). Data integration: A logic-based perspective. *AI Magazine*, 26(1):59–70.
- Ciociou, M. and Nau, D. (2000). Ontology-based semantics. In Cohn, A. G., Giunchiglia, F., and Selman, B., editors, *KR 2000, Principles of Knowledge Representation and Reasoning, Proceedings of the Seventh International Conference, Breckenridge, Colorado, USA, April 11-15, 2000*, pages 539–548. Morgan Kaufmann.
- Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Reddivari, P., Doshi, V., and Sachs, J. (2004). Swoogle: a search metadata engine for the semantic web. In Evans, D. A., Gravano, L., Herzog, O., Zhai, C., and Ronthaler, M., editors, *CIKM 2004. Proceedings of the Thirteenth ACM Conference on Information & Knowledge Management. Washington, DC, USA, November 8–13, 2004*, pages 652–659. ACM Press.
- Ding, Y. and Foo, S. (2002). Ontology research and development. Part 2 — a review of ontology mapping and evolving. *Journal of Information Science*, 28(5):383–396.
- Doan, A., Halevy, A., and Noy, N. F. (2004a). Semantic Integration Workshop at the 2nd International Semantic Web Conference (ISWC-2003). *SIGMOD Record*, 33(1):138–140.
- Doan, A., Noy, N. F., and Halevy, A. (2004b). Introduction to the special issue on semantic integration. *SIGMOD Record*, 33(4):11–13.
- Enderton, H. (2002). *A Mathematical Introduction to Logic*. Academic Press, second edition.
- Farquhar, A., Fikes, R., and Rice, J. (1997). The Ontolingua Server: a tool for collaborative ontology construction. *International Journal of Human-Computer Studies*, 46(6):707–727.
- Ganter, B. and Wille, R. (1999). *Formal Concept Analysis*. Springer.
- Giunchiglia, F., Marchese, M., and Zaihrayeu, I. (2006). Encoding classifications into lightweight ontologies. In Sure, Y. and Domingue, J., editors, *The Semantic Web: Research and Applications. 3rd European Semantic Web Conference, ESWC 2006. Budva, Montenegro, June 11-14, 2006. Proceedings*, volume 4011 of *Lecture Notes in Computer Science*, pages 80–94. Springer.
- Goguen, J. (1991). A categorical manifesto. *Mathematical Structures in Computer Science*, 1(49–67).
- Goguen, J. (2005). Data, schema, ontology and logic integration. *Logic Journal of the IGPL*, 13(6):685–715.
- Goguen, J. and Burstall, R. (1992). Institutions: Abstract model theory for specification and programming. *Journal of the ACM*, 39(1):95–146.
- Goguen, J. and Roşu, G. (2002). Institution morphisms. *Formal Aspects of Computing*, 13:204–307.
- Gruber, T. (1993). A translation approach for portable ontology specifications. *Knowledge Engineering*, 5(2):199–220.
- Grüninger, M. (2005). Model-theoretic approaches to semantic integration. In Kalfoglou et al. (2005).
- Grüninger, M. and Koppena, J. B. (2005). Semantic integration through invariants. *AI Magazine*, 26(1):11–20.
- Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, 43:625–640.
- Guarino, N. (1998a). Formal ontology and information systems. In Guarino (1998b), pages 3–15.
- Guarino, N., editor (1998b). *Formal Ontology in Information Systems. Proceedings of the First International Conference (FOIS 1998), Trento, Italy, June 6–8, 1998*, volume 46 of *Frontiers in Artificial Intelligence and Applications*. IOS Press.
- Guarino, N., Carrara, M., and Giaretta, P. (1994). Formalizing ontological commitments. In *Proceedings of the Twelfth National Conference on Artificial Intelligence AAAI-94, Seattle, Washington, USA, July 31–August 4, 1994*, pages 560–567. AAAI Press.
- Guarino, N. and Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological clarification. In Mars, N. J. I., editor, *Towards Large Knowledge Bases: Knowledge Building & Knowledge Sharing 1995. Proceedings of the International Conference on Building and Sharing Very Large-Scale Knowledge Bases, Enschede, The Netherlands, April 10–13, 1995*, pages 25–32. IOS Press.

- Hitzler, P., Krötzsch, M., Ehrig, M., and Sure, Y. (2005). What is ontology merging? — A category-theoretic perspective using pushouts. In *AAAI'05 Workshop on Contexts and Ontologies: Theory, Practice and Applications*. Pittsburgh, Pennsylvania, USA, July 9, 2005.
- Jannink, J., Pichai, S., Verheijen, D., and Wiederhold, G. (1998). Encapsulation and composition of ontologies. In *AAAI'98 Workshop on AI and Information Integration*, Madison, Wisconsin, USA.
- Kalfoglou, Y., Alani, H., Schorlemmer, M., and Walton, C. (2004). On the emergent semantic web and overlooked issues. In McIlraith, S. A., Plexousakis, D., and van Harmelen, F., editors, *The Semantic Web — ISWC 2004. Third International Semantic Web Conference. Hiroshima, Japan, November 7–11, 2004. Proceedings*, volume 3298 of *Lecture Notes in Computer Science*, pages 576–590. Springer.
- Kalfoglou, Y. and Schorlemmer, M. (2003a). IF-Map: An ontology-mapping method based on information-flow theory. In Spaccapietra, S., March, S., and Aberer, K., editors, *Journal on Data Semantics I*, volume 2800 of *Lecture Notes in Computer Science*, pages 98–127. Springer.
- Kalfoglou, Y. and Schorlemmer, M. (2003b). Ontology mapping: The state of the art. *The Knowledge Engineering Review*, 18(1):1–31.
- Kalfoglou, Y. and Schorlemmer, M. (2004). Formal support for representing and automating semantic interoperability. In Bussler, C., Davies, J., Fensel, D., and Studer, R., editors, *The Semantic Web: Research and Applications. First European Semantic Web Symposium, ESWS 2004. Heraklion, Crete, Greece, May 10-12, 2004. Proceedings*, volume 3053 of *Lecture Notes in Computer Science*, pages 45–60. Springer.
- Kalfoglou, Y., Schorlemmer, M., Sheth, A., Staab, S., and Uschold, M., editors (2005). *Semantic Interoperability and Integration*, volume 04391 of *Dagstuhl Seminar Proceedings*. IBFI, Schloss Dagstuhl, Germany.
- Kent, R. E. (2000). The information flow foundation for conceptual knowledge organization. In *6th International Conference of the International Society for Knowledge Organization*, Toronto, Canada.
- Kent, R. E. (2005). Semantic integration in the information flow framework. In Kalfoglou et al. (2005).
- Lassila, O. and Swick, R. (1999). Resource description framework (RDF) model and syntax specification. Technical report, World Wide Web Consortium (W3C). <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.
- Lucanu, D., Li, Y. F., and Dong, J. S. (2006). Semantic web languages — towards an institutional perspective. In Futatsugi, K., Jouannaud, J.-P., and Meseguer, J., editors, *Algebra, Meaning, and Computation. Essays Dedicated to Joseph A. Goguen on the Occasion of His 65th Birthday*, volume 4060 of *Lecture Notes in Computer Science*, pages 99–123. Springer.
- Lüttich, K. and Mossakowski, T. (2004). Specification of ontologies in CASL. In Varzi, A. C. and Vieu, L., editors, *Formal Ontology in Information Systems. Proceedings of the Third International Conference (FOIS 2004), Torino, Italy, November 4–6, 2004*, volume 114 of *Frontiers in Artificial Intelligence and Applications*, pages 140–150. IOS Press.
- Lüttich, K., Mossakowski, T., and Krieg-Brückner, B. (2004). Ontologies for the semantic web in CASL. In Fiadeiro, J. L., Mosses, P. D., and Orejas, F., editors, *Recent Trends in Algebraic Development Techniques. 17th International Workshop, WADT 2004, Barcelona, Spain, March 27-29, 2004, Revised Selected Papers*, volume 3423 of *Lecture Notes in Computer Science*, pages 106–125. Springer.
- Mac Lane, S. (1998). *Categories for the Working Mathematician*. Springer, second edition.
- McGuinness, D. and van Harmelen, F. (2004). OWL web ontology language. Technical report, World Wide Web Consortium (W3C). <http://www.w3.org/TR/2004/REC-owl-features-20040210/>.
- Menzel, C. (2002). Ontology theory. In Euzenat, J., Gómez-Pérez, A., Guarino, N., and Stuckenschmidt, H., editors, *Ontologies and Semantic Interoperability*, number 64 in CEUR Workshop Proceedings.
- Menzel, C. (2005). Basic semantic integration. In Kalfoglou et al. (2005).
- Meseguer, J. (1989). General logics. In Ebbinghaus, H.-D., Fernández-Prida, J., Garrido, M., Lascar, D., and Rodríguez-Artalejo, M., editors, *Logic Colloquium '87. Proceedings of the Colloquium Held in Granada, Spain, July 20–25, 1987*, volume 129 of *Studies in Logic and the Foundations of Mathematics*, pages 275–329. Elsevier Science.
- Mossakowski, T., Maeder, C., and Lüttich, K. (2007). The Heterogeneous Tool Set. In Grumberg, O. and Huth, M., editors, *Tools and Algorithms for the Construction and Analysis of Systems. 13th International Conference, TACAS 2007, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2007 Braga, Portugal, March 24 - April 1, 2007. Proceedings*, volume 4424 of *Lecture Notes in Computer Science*, pages 519–522. Springer.
- Mosses, P. D., editor (2004). *CASL Reference Manual: The Complete Documentation of the Common Algebraic Specification Language*, volume 2960 of *Lecture Notes in Computer Science*. Springer.
- Noy, N. F. (2004). Semantic integration: a survey of ontology-based approaches. *SIGMOD Record*, 33(4):65–69.
- Noy, N. F., Doan, A., and Halevy, A. Y. (2005). Semantic integration. *AI Magazine*, 26(1):7–9.
- Pierce, B. (1991). *Basic Category Theory for Computer Scientists*. MIT Press.
- Rahm, E. and Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. *The VLDB Journal*, 10(4):334–350.
- Santini, S. (2006). Summa contra ontologiam. In Grust, T., Höpfner, H., Illarramendi, A., Jablonski, S., Mesiti, M., Müller, S., Patranjan, P.-L., Sattler, K.-U., Spiliopoulou, M., and Wijsen, J., editors, *Current Trends in Database Technology – EDBT 2006. EDBT 2006 Workshops PhD, DataX, IIDB, IIHA, ICSNW, QLQP, PIM, PaRMA, and Reactivity on the Web, Munich, Germany, March 26-31, 2006, Revised Selected Papers*, volume 4254 of *Lecture Notes in Computer Science*, pages 483–496. Springer.
- Schorlemmer, M. and Kalfoglou, Y. (2003). On semantic interoperability and the flow of information. In Doan, A., Halevy, A., and Noy, N., editors, *Semantic Integration*, volume 82 of *CEUR Workshop Proceedings*.
- Schorlemmer, M. and Kalfoglou, Y. (2005). Progressive ontology alignment for meaning coordination: An information-theoretic foundation. In Dignum, F. et al., editors, *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, Utrecht, The Netherlands, July 25–29, 2005*, pages 737–744. ACM Press.
- Schorlemmer, M., Kalfoglou, Y., and Atencia, M. (2007). A formal foundation for ontology-alignment interaction models. *International Journal on Semantic Web and Information Systems*, 3(2):50–68.

- Shvaiko, P. and Euzenat, J. (2005). A survey of schema-based matching approaches. In Spaccapietra, S., March, S., and Aberer, K., editors, *Journal on Data Semantics IV*, volume 3730 of *Lecture Notes in Computer Science*, pages 146–171. Springer.
- Shvaiko, P. and Euzenat, J. (2007). Special issue: Ontology matching. *International Journal on Semantic Web and Information Systems*, 3(2):i–iii.
- Smith, B. (2003). Ontology. In Floridi, L., editor, *Blackwell Guide to the Philosophy of Computing and Information*, pages 155–166. Blackwell.
- Stumme, G. and Maedche, A. (2001). FCA-Merge: Bottom-up merging of ontologies. In *17th International Joint Conference on Artificial Intelligence (IJCAI'01)*, Seattle, Washington, USA.
- Uschold, M., Healy, M., Williamson, K., Clark, P., and Woods, S. (1998). Ontology reuse and application. In Guarino (1998b), pages 179–192.
- Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., and Hübner, S. (2001). Ontology-based integration of information — a survey of existing approaches. In *IJCAI 2001 Workshop on Ontologies and Information Sharing, Seattle, Washington, USA*.
- Zimmermann, A., Krötzsch, M., Euzenat, J., and Hitzler, P. (2006). Formalizing ontology alignment and its operations with category theory. In Bennett, B. and Fellbaum, C., editors, *Formal Ontology in Information Systems. Proceedings of the Fourth International Conference (FOIS 2006), Baltimore, Maryland, USA, November 9–11, 2006*, volume 150 of *Frontiers in Artificial Intelligence and Applications*, pages 277–288. IOS Press.