# A Formal Foundation for
# Ontology-Alignment Interaction Models

Marco Schorlemmer[1]        Yannis Kalfoglou[2]        Manuel Atencia[1]


[1]IIIA, Artificial Intelligence Research Institute
CSIC, Spanish National Research Council
Bellaterra (Barcelona), Catalonia, Spain

[2]AKT, Advanced Knowledge Technologies
School of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom

**Abstract.** Ontology alignment foundations are hard to find in the literature. The abstract nature of the topic and the diverse means of practice make it difficult to capture it in a universal formal foundation. We argue that such a lack of formality hinders further development and convergence of practices, and in particular, prevents us from achieving greater levels of automation. In this article we present a formal foundation for ontology alignment that is based on interaction models between heterogeneous agents on the Semantic Web. We use the mathematical notion of information flow in a distributed system to ground our three hypotheses of enabling semantic interoperability and we use a motivating example throughout the article: how to progressively align two ontologies of research quality assessment through meaning coordination. We conclude the article with the presentation—in an executable specification language—of such an ontology-alignment interaction model.

## 1. Introduction

Semantic heterogeneity is a phenomenon that emerges mostly in distributed heterogeneous environments, and it is addressed by a wide variety of communities and through the application of many diverse technologies. Its roots date back to the early stages of federated databases (Sheth & Larson, 1990) and has been continuously under investigation by database researchers through the application of a variety of (semi-)automatic schema matching techniques like those listed in (Rahm & Bernstein, 2001). It is well known that, for two separate systems to be capable of interoperating, exchanging vocabulary and syntax is insufficient because one needs also to agree upon the meaning of the communicated syntactic constructs. Separate systems, though, are most often engineered assuming different, sometimes even incompatible, conceptualisations. Ontologies have been advocated as a solution to this semantic heterogeneity: separate systems would need to match their own conceptualisations against a common ontology of the application domain, so that all communication is done according to the constraints derived from the ontology.

Although the use of ontologies may indeed favour semantic interoperability, it relies on the existence of agreed domain ontologies in the first place. Furthermore, these ontologies will have to be as complete and as stable for a domain as possible, because different versions only introduce more semantic heterogeneity. The use of ontologies for semantic integration is more in tune with a classical codification-centred knowledge management tradition, as put forward by Corrêa da Silva and Agustí (2003). Such tradition comprises the efforts to define standard upper-level ontologies such as CyC (Lenat, 1995) and SUO (Standard Upper Ontology Working Group, 2003), or to establish public ontology repositories for specific domains to favour knowledge reuse such as the Ontolingua server (Farquhar, Fikes, & Rice, 1997). Corrêa da Silva and Agustí (2003) remark that "centralised ontologies... promise to bring the control of the organization back to what was possible under classical management techniques. The problem is that they may also bring back the rigidity of agencies organized under the classical management tenets" (p. 130). Thus, semantic-integration approaches based on *a priori* common domain ontologies may be useful for clearly delimited and stable domains, but they are untenable and even undesirable in highly distributed, open, and dynamic environments such as the Semantic Web.

As a result, when ontology engineers begun to apply their products to the Semantic Web with the aim of solving the semantic heterogeneity problem, it became apparent that it would yield a new form of heterogeneity: that of ontology heterogeneity. The problem currently attracts the attention of practitioners

with different backgrounds and perspectives, ranging from the database community (Doan & Halevy, 2005) to researchers investigating ontology mapping and matching approaches (Kalfoglou & Schorlemmer, 2003b; Noy, 2004; Shvaiko & Euzenat, 2005). Despite the plethora of potential solutions, however, there are issues that still remain unclear and for which the researchers do not share a universal understanding. Part of the problem is due to the fact that the majority of work in ontology mapping or database schema matching is based on techniques that use syntactic and structural features of ontologies. The emphasis is on automation, scalability and (re-)use of alignment algorithms but there is an apparent lack of formal foundations for most of this work (with the notable exceptions of (Alagic & Bernstein, 2002) and (Bench-Capon, Malcolm, & Shave, 2003), for example). Even when formal foundations and theory take front stage, like in (Kent, 2005), there is a lack of practical implementations that provide insight to the application of the proposed theory.

The quest is to find the right balance between theory and practice on one hand, and to bridge the gap between syntax-based and semantic-based solutions on the other. It seems that these core topics are antithetical: the more practical an approach is, the less semantically rich it is; the more syntax-oriented an approach is, the less practical it will be—arguably, since semantically rich approaches are the Holy Grail of semantic integration. This illustration of the problem is probably not applicable to all situations. Notable exceptions like (Giunchiglia & Shvaiko, 2003)—with a rich semantic flavour—claim to be practical, but they still represent only a fraction of reported systems (see, for example the overviews in (Kalfoglou & Schorlemmer, 2003b; Noy, 2004; Shvaiko & Euzenat, 2005)). Still, as we strive for semantically rich and practical approaches we need to find the right theoretical foundation that will support them.

The angle from which we approach this is by looking at the mathematics of information flow underlying semantic alignment. More specifically, we are interested in addressing the progressive alignment of ontologies. In open, distributed, and highly decentralised environments it is more realistic to think of progressively achieving certain levels of semantic interoperability by coordinating and negotiating the meaning attached to syntactic constructs on the fly. For this reason, we do not want to focus on a particular ontology matching technique that, following a classical functional approach, takes mismatching ontologies as input and produces a suitable semantic alignment as output. Instead we want to shift our attention to what basic capability an agent should have to be able to engage in an ontology-alignment interaction when required. Schorlemmer and Kalfoglou (2005) gave a channel-theoretic foundation for meaning coordination when two agents align their respective local concepts by exchanging instances of these concepts. In this article we elaborate on this idea and formalise semantic alignment as a process of information-channel refinement. We also take the framework resulting from this formalisation to specify a basic and general ontology-alignment process that translates directly into an executable interaction model when grounded on particular agents.

The article is structured as follows. In Section 2 we frame our approach to the ontology alignment as a process of meaning coordination and state our basic assumptions. Next, in Section 3 we give three hypotheses that guide our formalisation of semantic alignment and we introduce the necessary mathematical preliminaries. Section 4 is the core of the article, where we model ontology alignment as a process of information-channel refinement. In Section 5 we then describe this process from an operational perspective and give an executable specification of it as an agent interaction model. Section 6 concludes the article.

## 2. Aligning Ontologies through Meaning Coordination

Before deploying our formal foundation to address the ontology alignment problem, we first frame the problem by stating the assumptions upon which we shall later build an interaction model for ontology alignment.

We shall consider a scenario in which two agents $A_1$ and $A_2$ want to interoperate, but each agent $A_i$ has its knowledge represented according to its own conceptualisation, which we assume is explicitly specified by means of its own ontology $O_i$. Any expression $\alpha_i$ using the vocabulary $O_i$ will be considered semantically distinct *a priori* from any expression $\alpha_j$ using vocabulary $O_j$ (with $j \neq i$), even if they happen to be syntactically equal, unless the semantic evidence unveiled by an ontology-alignment process of the kind described below makes them mean the same to $A_1$ and $A_2$. Furthermore, we assume that the agents' ontologies are not open to other agents for inspection, so that semantic heterogeneity cannot be solved by semantically matching the ontologies beforehand.

An agent may learn about the ontology of another agent only through meaning coordination. Thus, we assume that agent $A_i$ is capable of requesting from agent $A_j$ to explain the intended meaning of an expression $\alpha_j$ that is in a message from $A_j$ to $A_i$ and uses the vocabulary $O_j$. Agent $A_i$ might request such

an explanation with the intention of determining the semantic relationship of the fragment of $O_j$ used in $\alpha_j$ with respect to its local vocabulary $O_i$. Correspondingly, we assume that agent $A_j$ is capable of explaining $A_i$ the meaning of expression $\alpha_j$ by means of a *token* of this expression.

The formal framework we describe in this article is neutral with respect to the syntactic form of expressions and, more importantly, to what tokens might be, giving an interesting level of generality to the ontology-alignment interaction models discussed below. The Oxford Dictionary of English defines a token as "a thing serving as a visible or tangible representation of something abstract." In our scenario a token will be something agent $A_i$ is capable of processing and putting into relationship with its own local ontology $O_i$.

Take for instance the ontology negotiation process described by Bailin and Truszkowski (2002). There, an agent $A_i$, upon the reception from another agent $A_j$ of a message containing a list of keywords, either sends to $A_j$ an interpretation of the keywords in the form of WordNet synonyms in order to check that it has interpreted $A_j$'s vocabulary correctly, or else requests $A_j$ for a clarification of the interpretation of unknown keywords, also in form of WordNet synonyms. Thus, in this scenario, the role of tokens is played by WordNet synonyms of those keywords whose interpretation needs to be confirmed or clarified.

Looking at another ontology alignment scenario, Wang and Gasser (2002) present an ontology-alignment algorithm for open multi-agent systems, where ontologies are partitions of domain instances into categories, based on the K-means algorithm, a typical partition-based clustering method. The alignment is carried out in an online fashion by exchanging instances between two agents, rather than by exchanging abstract concepts. When an agent plans to express some concept or category to other agents it uses an instance belonging to that category to represent this concept. In this scenario it is particular domain instances that play the role of tokens of a concept or category. Wang and Gasser further note that, "unless a set of agents *already has* a compatible and verified shared ontology, it is difficult to see how they could specify categories to each other in another way." The capability of a set of agents to process and classify tokens according to their own local ontologies is what underlies the ontology-alignment process. Van Diggelen, Beun, Dignum, van Eijk, and Meyer (2006) describe also an ontology alignment protocol based on pointing to instances for concept explication. One agent communicates a number of positive and negative examples of the concept to the other agent, which classifies these examples using the concept classifier from its own ontology.

Finally, in yet other scenarios, Giunchiglia and Shvaiko (2004) and Bouquet, Serafini, and Zanobini (2003) use mappings of concepts in a tree hierarchy to propositional expressions using WordNet synsets in order to check, by means of a SAT prover, the semantic relationships between concepts occurring in two different hierarchies. In this scenario, a concept is represented by a propositional formula, playing the role of the token for this concept, which can then be processed by each agent with the SAT prover.

In this paper we shall use as running example the following realistic scenario involving the research quality assessment of a researcher's publication record:

*Example (Research Quality Assessment)*. It is common to qualitatively classify journal publications in accordance to standardised and consensual classification schemes. For example, Thompson Scientific (the former Institute of Scientific Information – ISI) offers classifications of scholarly journals according to their *impact factor*. On the other hand, popular citation repositories like CiteSeer, use different classification schemes based on the *number of citations*. Let us assume that in a research assessment exercise (similar to UK's Research Assessment Exercise series of quality assessment or to those conducted by the Quality Assurance Agency for the University System in Catalonia) an individual researcher and a quality assessment agency need to interoperate for the benefit of exchanging up to date information regarding the quality of publications.

Let agent $A_1$ (a computer-science researcher) conceptualise the quality of journal articles in his or her CV on the basis of the journal's citation ratio. Let it use for this two distinctive categories: high-citation-ratio journals, which are those that belong to the top 25% of the CiteSeer list of estimated impact of publication venues in computer science (CiteSeer, 2003); low-citation-ratio journals that belong to the bottom 20% of that list. Let agent $A_2$ (a research quality assessment agency) use a conceptualisation based on Thompson Scientific's impact factor (Garfield, 1994). For computer science Thompson Scientific indexes 352 journals, which can be ranked according to their impact factor (Journal Citation Reports, 2005). Let agent $A_2$ also use two distinctive categories, high-impact journals and low-impact journals. The first ones are those that have an impact factor greater than 1, whereas journals with impact factors lower than 0.5 are regarded as low-impact. Agent $A_1$ may have received a request from $A_2$ to provide a list of its high-impact papers, terminology $A_1$ needs to align with high-citation-ratio vs. low-citation-ratio papers in $O_1$. For this reason $A_1$ may request a token representing the meaning of high-impact, for which $A_2$ uses a journal's ISSN. Reciprocally, $A_1$ may explain to $A_2$ the meaning of high-citation-ratio (or low-citation-ratio) using an

ISSN as token, too. Both $A_1$ and $A_2$ are capable of processing tokens (ISSN's) and of classifying them according to their own ontologies. In order to do this, $A_1$ searches in CiteSeer's data (using the journal name it gets from the ISSN), whereas $A_2$ consults Thomson Scientific's Journal Citation Reports using the ISSN directly.


## 3. The Formal Framework

We have described ontology alignment as a process by which agents make the intended meaning of syntactic expressions explicit to each other through the use of tokens for these expressions. We deliberately have left unspecified what these tokens actually are, and have only briefly mentioned that we shall consider tokens as something agents are capable of processing and putting into relationship with their own local vocabulary. This view of ontology alignment is the result of the research initiated by Schorlemmer and Kalfoglou in (2003) and Kalfoglou and Schorlemmer (2004) aiming at a formal foundation for semantic interoperability and integration based on channel theory—Barwise and Seligman's proposal for a mathematical theory of information (Barwise & Seligman, 1997).

The original effort to develop an information-theoretic approach to ontology alignment using Barwise and Seligman's theory around the issues of organising and relating ontologies is Kent's Information Flow Framework (IFF) (Kent, 2000). Recently, Kent has proposed a formal characterisation of semantic integration in terms of IFF (Kent, 2005). Also recently, Goguen has shown that Barwise and Seligman's theory can be expressed in terms of institution theory (Goguen & Burstall, 1992), and he uses this insight to provide foundations for principled semantic integration (Goguen, in press).

In this section we introduce the main channel-theoretic constructs required for our formal foundation for ontology alignment, motivating them by means of three *Semantic-Alignment Hypotheses*.

**Hypothesis 1:** *Semantic alignment is relative to both the syntactic expressions to be aligned, which we shall call types, and the entities used as tokens for these types. This context, which is local to an agent, can be naturally modelled by means of a classification (Definition 1).*

Channel theory takes the idea of a classification as the fundamental notion for modelling the local context by which tokens relate to types:

**Definition 1:** A classification $\mathbf{A} = \langle tok(\mathbf{A}), typ(\mathbf{A}), \models_{\mathbf{A}} \rangle$ consists of a set of tokens $tok(\mathbf{A})$, a set of types $typ(\mathbf{A})$ and a classification relation $\models_{\mathbf{A}} \subseteq tok(\mathbf{A}) \times typ(\mathbf{A})$ that classifies tokens to types.

Although a very simple notion, classifications have recently been used, under varying terminology, in many related fields of formal knowledge representation and theoretical computer science (e.g., in algebraic logic (Dunn & Hardegree, 2001), categorical logic (Barr, 1996), formal concept analysis (Ganter & Wille, 1999), and process algebra (Pratt, 2001)).

**Hypothesis 2:** *Semantic alignment presupposes a flow of information between expressions (i.e., types) of separate agents that happens by virtue of shared tokens for these expressions. This flow of information can be accurately described by means of an information channel (Definition 3).*

A fundamental construct of channel theory is that of an information channel between two classifications. It models the information flow between components. First, though, we need to describe how classifications are connected with each other through infomorphisms:

**Definition 2:** An infomorphism $f = \langle f^{\rightarrow}, f^{\leftarrow} \rangle : \mathbf{A} \rightarrow \mathbf{B}$ from classifications $\mathbf{A}$ to $\mathbf{B}$ is a contra-variant pair of functions $f^{\rightarrow} : typ(\mathbf{A}) \rightarrow typ(\mathbf{B})$ and $f^{\leftarrow} : tok(\mathbf{B}) \rightarrow tok(\mathbf{A})$ satisfying the following fundamental property, for each type $\alpha \in typ(\mathbf{A})$ and token $b \in tok(\mathbf{B})$:

$$f^{\leftarrow}(b) \models_{\mathbf{A}} \alpha \quad \text{iff} \quad b \models_{\mathbf{B}} f^{\rightarrow}(\alpha)$$

$$
\begin{array}{ccc}
\alpha & \xmapsto{\ f^{\rightarrow}\ } & f^{\rightarrow}(\alpha) \\
\Big\vert \models_{\mathbf{A}} & & \Big\vert \models_{\mathbf{B}} \\
f^{\leftarrow}(b) & \xleftarrow{\ f^{\leftarrow}\ } & b
\end{array}
$$

As with classifications, infomorphisms have been around in the literature for quite a time, and its contra-variance between the type- and token-level is recurrent in many fields. They would correspond to *interpretations* when translating between logical languages (Enderton, 2002), or to *Chu transforms* in the context of *Chu spaces* (Pratt, 1995). Channel theory makes use of this contra-variance to model the flow of information at type-level because of the particular connections that happen at the token-level:

**Definition 3:** An information channel consists of two classifications $\mathbf{A}_1$ and $\mathbf{A}_2$ connected through a core classification $\mathbf{C}$ via two infomorphisms $f_1$ and $f_2$:
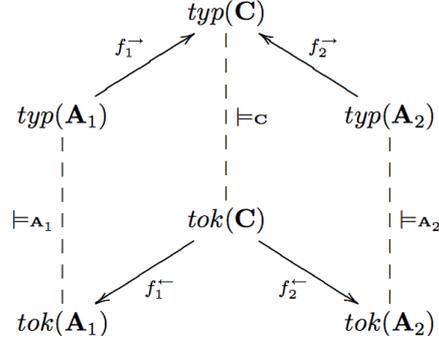


**Hypothesis 3:** *Semantic alignment is formally characterised by a consequence relation between expressions (i.e., types) of separate agents. This consequence relation can be faithfully captured by the natural logic (Definition 6) of the core of the information channel underlying the integration.*

Channel theory is based on the understanding that information flow is the result of regularities in distributed systems. These regularities are implicit in the representation of systems as interconnected classifications. However, one can make these regularities explicit in a logical fashion by means of theories and local logics:

**Definition 4:** A theory $T = \langle typ(T), \vdash_T \rangle$, consists of a set $typ(T)$ of types, and a binary relation $\vdash_T$ between subsets of $typ(T)$. Pairs $\langle \Gamma, \Delta \rangle$ of subsets of $typ(T)$ are called sequents. If $\Gamma \vdash_T \Delta$, for $\Gamma, \Delta \subseteq typ(T)$, then the sequent $\Gamma \vdash_T \Delta$ is called a constraint. T is regular if for all $\alpha \in typ(T)$ and all $\Gamma, \Gamma', \Delta, \Delta'$, $\Sigma \subseteq typ(T)$:
   1. Identity: $\alpha \vdash_T \alpha$.
   2. Weakening: If $\Gamma \vdash_T \Delta$, then $\Gamma, \Gamma' \vdash_T \Delta, \Delta'$.
   3. Global Cut: If $\Gamma, \Sigma_0 \vdash_T \Delta, \Sigma_1$ for each partition $\langle \Sigma_0, \Sigma_1 \rangle$ of $\Sigma$, then $\Gamma \vdash_T \Delta$.

Note that, at it is usual with sequents and constraints, we write $\alpha$ instead of $\{\alpha\}$ and $\Gamma, \Gamma'$ instead of $\Gamma \quad \Gamma'$. Also, a partition of $\Sigma$ is a pair $\langle \Sigma_0, \Sigma_1 \rangle$ of subsets of $\Sigma$, such that $\Sigma_0 \cup \Sigma_1 = \Sigma$ and $\Sigma_0 \cap \Sigma_1 = \varnothing$; $\Sigma_0$ and $\Sigma_1$ may themselves be empty (hence it is actually a quasi-partition). Note that Global Cut is implied by the usual (Finitary) Cut only if the binary relation is *compact*, i.e., $\Gamma \vdash_T \Delta$ implies the existence of finite subsets $\Gamma_0 \subseteq \Gamma$ and $\Delta_0 \subseteq \Delta$ such that $\Gamma_0 \vdash_T \Delta_0$.

Regularity arises from the observation that, given any classification of tokens to types, the set of all sequents that are satisfied by all tokens always fulfills Identity, Weakening, and Global Cut. Hence, the notion of a local logic:

**Definition 5:** A local logic $\mathcal{L} = \langle tok(\mathcal{L}), typ(\mathcal{L}), \models_{\mathcal{L}}, \vdash_{\mathcal{L}}, N_{\mathcal{L}} \rangle$ consists of a classification $cla(\mathcal{L}) = \langle tok(\mathcal{L}), typ(\mathcal{L}), \models_{\mathcal{L}} \rangle$, a regular theory $th(\mathcal{L}) = \langle typ(\mathcal{L}), \vdash_{\mathcal{L}} \rangle$ and a subset of $N_{\mathcal{L}} \subseteq tok(\mathcal{L})$ of normal tokens, which satisfy all the constraints of $th(\mathcal{L})$; a token $a \in tok(\mathcal{L})$ satisfies a constraint $\Gamma \vdash_{\mathcal{L}} \Delta$ of $th(\mathcal{L})$ if, when $a$ is of all types in $\Gamma$, $a$ is of some type in $\Delta$.

Finally, every classification determines a *natural logic*, which captures the regularities of the classification in a logical fashion, and which we shall use in order to model the semantic interoperability between agents with different ontologies:

**Definition 6:** The natural logic of a classification $\mathbf{C}$ is the local logic $Log(\mathbf{C})$ generated from $\mathbf{C}$, and has as classification $\mathbf{C}$, as regular theory the theory whose constraints are the sequents satisfied by all tokens,

and whose tokens are all normal.


## 4. Ontology Alignment as Information-Channel Refinement

Recall that we are assuming that agents explain each other the intended meaning of a syntactic expression $\alpha$ by exchanging a token a that serves as a representation of $\alpha$. Consequently, we take a classifying assertion of the form "token $a$ is of type $\alpha$" to be our basic *unit of meaning coordination*. To make this sort of meaning explanation feasible we need each agent to be capable of processing tokens that are exchanged during an alignment process and of putting them into relationship with its own local ontology. We shall denote with $U$ this set of tokens. Alignment is, therefore, relative to the set $U$. In our Research Quality Assessment example, $U$ is the set of ISSN's, or a subset of it. It could, for instance, be the set of ISSN's of those journals that agent $A_1$ has published in.

Following the three Semantic-Alignment Hypotheses of Section 3 as our guide, we shall show that this alignment process determines a sequence of information-channel refinements. According to Hypothesis 1, there will be classifications that naturally model the contexts in which the alignment of ontologies occurs, and we shall first show how these classifications arise. Next, by Hypothesis 2, an information channel should enable us to capture the alignment achieved when meaning is coordinated (based on our units of meaning coordination) between agents, and to do so by focusing on the tokens that have been exchanged. Finally, by Hypothesis 3, a consequence relation between expressions of agents will faithfully capture the semantic integration achieved through meaning coordination. We will show that each new unit of meaning coordination between agents actually leads to a consequence relation that more precisely formalises the information flow between agents because it will be determined by a more refined information channel.


### 4.1. Contexts of Meaning Coordination

For the purposes of meaning coordination described in this article, we adopt a definition of ontology that includes most of its core components: *concepts*, organised in an *is-a hierarchy* with notions of *disjointness* of two concepts—when no token can be considered of both concepts—and *coverage* of two concepts—when all tokens are covered by two concepts. Both disjointness and coverage can easily be extended to more than two concepts. Disjointness and coverage are typically specified by means of ontological axioms. In this article we take these kind of axioms into account including disjointness and coverage into the hierarchy of concepts by means of two binary relations '$\perp$' and '|', respectively. In (Kalfoglou & Schorlemmer, 2003a), we included also *roles* in their core treatment of ontologies. We have left them out here for the ease of presentation.

**Definition 7**: An ontology is a tuple $O = \langle C, \leq. \perp, | \rangle$ where $C$ is a finite set of concept symbols; $\leq$ is a reflexive, anti-symmetric and transitive relation on $C$ (a partial order); $\perp$ is an irreflexive and symmetric relation on $C$ called disjointness; | is a symmetric relation on $C$ called coverage; and the following conditions are satisfied:
    1. For all $c, c', d, d' \in C$, if $c \perp d, c' \leq c$ and $d' \leq d$ then $c' \perp d'$.
    2. For all $c, d, e \in C$, if $c \mid d$ and $d \leq e$ then $c \mid e$.

Our approach to meaning coordination uses the fact that, first, an ontology $O = \langle C, \leq. \perp, | \rangle$ always determines a regular theory $Th(O)$ by taking the smallest regular theory (i.e., the smallest theory closed under Identity, Weakening, and Global Cut) over $typ(Th(O)) = C$ such that, for all $c, d \in C$,

$$
\begin{aligned}
c \vdash_{Th(O)} d & \quad\text{iff}\quad & c \leq d \\
c, d \vdash_{Th(O)} & \quad\text{iff}\quad & c \perp d \\
\vdash_{Th(O)} c, d & \quad\text{iff}\quad & c \mid d
\end{aligned}
$$

Second, a regular theory $T$ can be represented as a classification $Cla(T)$ (Fundamental Representation Theorem (Barwise & Seligman, 1997)) by generating formally created tokens (called *formal tokens*) as follows:
    1. Let tokens be all those sequents $\langle \Gamma, \Delta \rangle$ that form a quasi-partition of the set of concepts ($\Gamma \cup \Delta = C$ and $\Gamma \cap \Delta = \varnothing$), and are not constraints of the theory ($\Gamma \vdash_T \Delta$).
    2. Classify these tokens according to the concepts that occur in the left-hand side component of the sequent.

*Example (Research Quality Assessment).* In our example, $O_1$ would be specified by $C_1$ = {high-citation-ratio, low-citation-ratio} with high-citation-ratio $\perp$ low-citation-ratio; $O_2$ would be specified by $C_2$ = {high-impact, low-impact} with high-impact $\perp$ low-impact. For the ontology of $O_1$ of our example, $Th(O_1)$ is the regular closure of

high-citation-ratio, low-citation-ratio $\vdash_{Th(O_1)}$

its formal tokens would be the sequents

$\langle$\{high-citation-ratio\}, \{low-citation-ratio\}$\rangle$
$\langle$\{low-citation-ratio\}, \{high-citation-ratio\}$\rangle$
$\langle\varnothing$, \{high-citation-ratio,low-citation-ratio\}$\rangle$

and the final classification would be

| $\models_{Cla(Th(O_1))}$ | high-citation-ratio | low-citation-ratio |
|---|---|---|
| $\langle$\{high-citation-ratio\}, \{low-citation-ratio\}$\rangle$ | 1 | 0 |
| $\langle$\{low-citation-ratio\}, \{high-citation-ratio\}$\rangle$ | 0 | 1 |
| $\langle\varnothing$, \{high-citation-ratio,low-citation-ratio\}$\rangle$ | 0 | 0 |

The generation of tokens by means of sequents and their classification to types may not seem obvious, but it is based on the fact that these sequents codify the content of the classification table (the left-hand sides of these sequents indicate which columns of the table bear a '1', while the right-hand sides indicate which columns bear a '0').

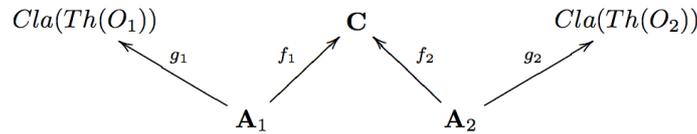## 4.2. Alignment through Meaning Coordination

We shall use a system of classifications and infomorphisms to model how $O_1$ and $O_2$ are progressively coordinated, capturing in turn the alignment achieved through meaning coordination between $A_1$ and $A_2$. Recall that we took statements of the form

"token $a$ is of type $\alpha$"

to be the agents' basic unit of meaning coordination. An alignment is therefore determined both at the type and at the token level, because
- $A_i$ will have attempted to explain $A_j$ a subset of its vocabulary, and
- it will have used tokens from $U$ for this purpose.

**Definition 8:** Given two ontologies $O_1$ and $O_2$ with set of concepts (i.e., types) $C_1$ and $C_2$, respectively, and a set of tokens $U$, an alignment of $O_1$ and $O_2$ with respect to $U$ is a system $\mathcal{A}$ of classifications and infomorphisms

$$Cla(Th(O_1)) \quad\xleftarrow{g_1}\quad \mathbf{C} \quad\xrightarrow{f_2}\quad Cla(Th(O_2))$$

$$\mathbf{A}_1 \xrightarrow{f_1} \quad \mathbf{A}_2 \xrightarrow{g_2}$$

where:
- $Cla(Th(O_i))$ are the classifications generated from the regular theory of ontologies $O_i$ (i.e., with formal tokens);
- $\mathbf{C}$ is the classification determined by the meaning coordination done so far, i.e, $tok(C) \subseteq U$ and $typ(\mathbf{C}) \subseteq C_1 \cup C_2$ are the set of those tokens and the *disjoint* union of types that the agents have used in their exchanges of units of meaning coordination, and $a \models_C \alpha$ if "token a is of type $\alpha$" has been an exchanged unit of meaning coordination (it is important to take the disjoint union, as $O_1$ and $O_2$ may contain some identical type symbols—we will subindex the types in this union in order to distinguish them);

- classifications $\mathbf{A}_i$ link types in $\mathbf{C}$ with their original types in $Cla(Th(O_i))$ and put tokens in $\mathbf{C}$ in relationship with the local types in $Cla(Th(O_i))$. They are defined as follows:
    - $typ(\mathbf{A}_i) \subseteq C_i$ are the subset of types of ontology $O_i$ used in exchanges of units of meaning coordination;
    - $tok(\mathbf{A}_i) = tok(Cla(Th(O_i)))$ are the set of formal tokens for agent $A_i$ ; and
    - $a \models_{\mathbf{A}_i} \alpha$ iff $a \models_{Cla(Th(O_i))} \alpha$.
- functions $g_i^{\rightarrow} : typ(\mathbf{A}_i) \to C_i$ are the type inclusion maps, and functions $g_i^{\leftarrow} : tok(Cla(Th(O_i))) \to tok(Cla(Th(O_i)))$ are identity maps on tokens.
- functions $f_i^{\rightarrow} : typ(\mathbf{A}_i) \to typ(\mathbf{C})$ are type injections into the disjoint union of types, and functions $f_i^{\leftarrow} : tok(\mathbf{C}) \to tok(Cla(Th(O_i)))$ model how each agent is putting the token $a \in U$ in relationship with each own ontology. They are defined as follows: $f_i^{\leftarrow}(a) = \langle \Gamma, \Delta \rangle$ iff, according to $A_i$, $a$ is token of all types in $\Gamma \subseteq C_i$ but of none of the types in $\Delta \subseteq C_i$. Recall that $\langle \Gamma, \Delta \rangle$ forms a (quasi-)partition of $C_i$.

Notice that the actual alignment is done through the information channel $\mathbf{A}_1 \overset{f_1}{\longrightarrow} \mathbf{C} \overset{f_2}{\longleftarrow} \mathbf{A}_2$.

*Example (Research Quality Assessment).* In our example, suppose the situation in which $A_1$ and $A_2$ have exchanged the following units of meaning coordination:

"token 1066-8888 is of type high-impact"
"token 1066-8888 is of type high-citation-ratio"

where 1066-8888 is the ISSN of the *VLDB Journal*. The alignment achieved so far would be modelled by a system of classifications and infomorphisms $\mathcal{A}$, in which $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{C}$ are given by classification tables

| $\models_{\mathbf{A}_1}$ | high-citation-ratio |
|---|---|
| $\langle\{\text{high-citation-ratio}\}, \{\text{low-citation-ratio}\}\rangle$ | 1 |
| $\langle\{\text{low-citation-ratio}\}, \{\text{high-citation-ratio}\}\rangle$ | 0 |
| $\langle\varnothing, \{\text{high-citation-ratio,low-citation-ratio}\}\rangle$ | 0 |

| $\models_{\mathbf{A}_2}$ | high-impact |
|---|---|
| $\langle\{\text{high-impact}\}, \{\text{low-impact}\}\rangle$ | 1 |
| $\langle\{\text{low-impact}\}, \{\text{high-impact}\}\rangle$ | 0 |
| $\langle\varnothing, \{\text{high-impact,low-impact}\}\rangle$ | 0 |

| $\models_{\mathbf{C}}$ | high-citation-ratio | low-impact |
|---|---|---|
| 1066-8888 | 1 | 1 |

and the functions $g_i^{\rightarrow}$ are type inclusions, $g_i^{\leftarrow}$ are token identities, and $f_i^{\rightarrow}$ are type injections. The semantic link is done at token level of information channel $\mathbf{A}_1 \overset{f_1}{\longrightarrow} \mathbf{C} \overset{f_2}{\longleftarrow} \mathbf{A}_2$; where

- $f_1^{\leftarrow}(1066\text{-}8888) = \langle\{\text{high-citation-ratio}\}, \{\text{low-citation-ratio}\}\rangle$ because 1066-8888 is the ISSN of the *VLDB Journal*, which is 51st in CiteSeer's list, in the top 4.17%. Agent $A_1$ considers it a journal with a high-citation ratio.
- $f_2^{\leftarrow}(1066\text{-}888) = \langle\{\text{high-impact}\}, \{\text{low-impact}\}\rangle$ because the *VLDB Journal* is regarded by agent $A_2$ as a high-impact journal as it has an impact factor of 4.317 and is 7th in Thomson Scientific's rank.
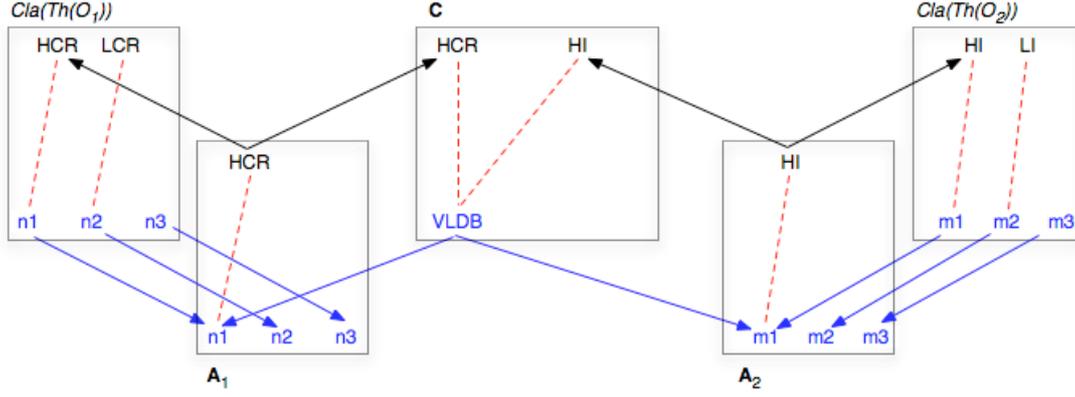
*Figure 1.* Partial alignment $\mathcal{A}$ in the Research Quality Assessment scenario.

Figure 1 shows the alignment $\mathcal{A}$ using line diagrams to illustrate classifications and infomorphisms. We use abbreviations HCR, LCR, HI, and LI for types high-citation-ratio, low-citation-ratio, high-impact, and low-impact; n1, n2, and n3 for the formally generated tokens of $Cla(Th(O_1))$; m1, m2, and m3 for the formally generated tokens of $Cla(Th(O_2))$; and VLDB for token 1066-8888.

## 4.3. Semantic Integration

We have seen that, in an alignment $\mathcal{A}$ as defined in Definition 8, the information channel $\mathbf{A}_1 \xrightarrow{\;f_1\;}$ $\mathbf{C} \xleftarrow{\;f_2\;} \mathbf{A}_2$ establishes the semantic link between types on the tokens used in units of meaning coordination. But the whole alignment $\mathcal{A}$ also determines an information channel between $Cla(Th(O_1))$ and $Cla(Th(O_2))$.

We know from (Barwise & Seligman, 1997) that, for every system of classifications and infomorphisms there exists a unique classification (up to isomorphisms), known as the *colimit* of he system, that captures the regularities established in each classification of the system and of the infomorphisms between them.

**Proposition 1:** Let $\mathcal{A}$ be an alignment as defined in Definition 8. Let $\mathbf{D}$ be the classification with $tok(\mathbf{D})$ $= tok(\mathbf{C})$, $typ(\mathbf{D}) = C_1 \cup C_2$, and classification relation $a \models_{\mathbf{D}} \alpha$ iff $a \models_{\mathbf{C}} \alpha$. Let $h_i : Cla(Th(O_1)) \to \mathbf{D}$ be the infomorphisms with $h_i^{\rightarrow}$ being the injection maps of the disjoint union, and $h_i^{\leftarrow} = f_i^{\leftarrow}$. Let $k : \mathbf{D} \to \mathbf{C}$ be the infomorphism with $k^{\rightarrow}$ being the type inclusion map, and $k^{\leftarrow}$ being the token identity map. Then $\mathbf{D}$, with infomorphisms $h_i$ and $k$, is a colimit of $\mathcal{A}$:
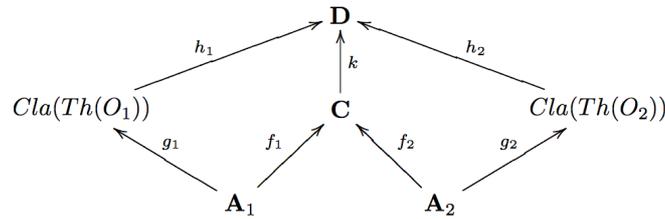


It is easy to proof that $\mathbf{D}$ is indeed a colimit of $\mathcal{A}$. Consequently, the natural logic of $\mathbf{D}$ models, by means of sequents with types in $C_1 \cup C_2$, the semantic integration achieved on the grounds of alignment $\mathcal{A}$. We shall call the information channel $Cla(Th(O_1)) \xrightarrow{\;h_1\;} \mathbf{D} \xleftarrow{\;h_2\;} Cla(Th(O_2))$ the semantic integration of $O_1$ and $O_2$ relative to $\mathcal{A}$.

*Example (Research Quality Assessment)*. In our example, the theory of the natural logic of the information channel's core determined by the colimit of the alignment shown in Figure 1 is the regular closure of the following set of constraints:

|– high-citation-ratio$_1$
|– high-impact$_2$

| | |
|---|---|
| low-citation-ratio$_1$ | I– |
| low-impact$_2$ | I– |

The representation of an ontology alignment as a system of objects and morphisms in a category (see Definition 8), and of semantic integration by means of a colimit of such a diagram, bears a close relationship to the notion of *W-alignment diagram* described in (Zimmermann, Krötzsch, Euzenat, & Hitzler, 2006). This is so because both notions share the same categorical approach to semantic alignment. But, unlike in (Zimmermann et al., 2006), we specifically take the type-token structure in semantic integration into account, and we define alignment with respect to this two-tier model. We claim that in this way we better capture Barwise and Seligman's basic insight that "information flow crucially involves both types and their particulars" (Barwise & Seligman, 1997, p. 27).

## 4.4. Refinement

Obviously, the natural logic of the core classification $\mathbf{D}$ will more accurately model the information flow between ontologies the more tokens we have in $\mathbf{D}$. The aim of meaning coordination would be to get more and more refined information channels that determine more reliable models of the semantic integration of ontologies. For this we use the idea of a channel refinement as given in (Barwise & Seligman, 1997):

**Definition 9:** Let $C$ be the information channel $\mathbf{A}_1 \overset{f_1}{\longrightarrow} \mathbf{C} \overset{f_2}{\longleftarrow} \mathbf{A}_2$ and let $C'$ be the information channel $\mathbf{A}_1 \overset{f'_1}{\longrightarrow} \mathbf{C'} \overset{f'_2}{\longleftarrow} \mathbf{A}_2$ between the same classifications $\mathbf{A}_1$ and $\mathbf{A}_2$. A refinement infomorphisms from $C'$ to $C$ is an infomorphism $r : \mathbf{C'} \to \mathbf{C}$ such that $f_i = r \circ f'_i$. We say that channel $C'$ is a refinement of $C$.

By successively exchanging units of meaning coordination, we define the sequence of alignments $\mathcal{A}^0, \mathcal{A}^1, \ldots, \mathcal{A}^n, \ldots$ as follows:

- initially, for $\mathcal{A}^0$, we have $typ(\mathbf{C}^0) = \varnothing$ and $tok(\mathbf{C}^0) = \varnothing$ (and thus $typ(\mathbf{A}_i^0) = \varnothing$);

- given alignment $\mathcal{A}^n$ (as the result of exchanging the $n$ units of meaning coordination), and exchanging the $(n + 1)$th unit of coordination "token $a$ is of type $\alpha$," we define alignment $\mathcal{A}^{n+1}$ to be:

$$typ(\mathbf{A}_i^{n+1}) \quad = \quad \begin{cases} typ(\mathbf{A}_i^n) \cup \{\alpha\} & \text{if } \alpha \in C_i \\ typ(\mathbf{A}_i^n) & \text{otherwise} \end{cases}$$

$$typ(\mathbf{C}^{n+1}) \quad = \quad typ(\mathbf{C}^n) \cup \{\alpha_k\} \quad \text{where } k = i \text{ if } \alpha \in C_i$$

$$tok(\mathbf{C}^{n+1}) \quad = \quad tok(\mathbf{C}^n) \cup \{a\}$$

$f_i^{n+1\leftarrow}$ extends $f_i^{n\leftarrow}$, with $f_i^{n+1\leftarrow}(a) = \langle \Gamma, \Delta \rangle$ if $a$ is classified by $A_i$ as of all types in $\Gamma$ but of no type in $\Delta$; and all $g_i^{n+1}$ and $f_i^{n+1}$ in alignment $\mathcal{A}^{n+1}$ are infomorphisms according to Definition 8. We write $\mathcal{A}^n \Rightarrow_{a|=\alpha} \mathcal{A}^{n+1}$.

The following proposition states that indeed exchanging units of meaning coordination refines the information channel that models the semantic integration of $O_1$ and $O_2$.

**Proposition 2 (Ontology Alignment as Information-Channel Refinement):** Let $\mathcal{A}^n \Rightarrow_{a|=\alpha} \mathcal{A}^{n+1}$, and let $\mathbf{D}^n$ and $\mathbf{D}^{n+1}$ be the colimits of $\mathcal{A}^n$ and $\mathcal{A}^{n+1}$ as given in Proposition 1, respectively. Then there exists an infomorphism $r : \mathbf{D}^{n+1} \to \mathbf{D}^n$ such that $r$ is a refinement infomorphism:
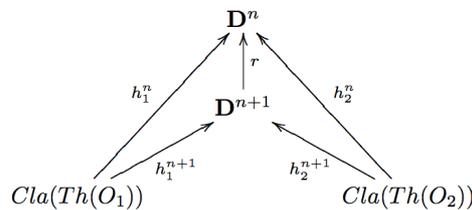
PROOF: By the way a sequence of alignments is defined, $tok(\mathbf{D}^n) \subseteq tok(\mathbf{D}^{n+1})$. Let $r^{\rightarrow}$ be the identity map and $r^{\leftarrow}$ be the inclusion map between tokens. It is trivial to see that $r$ is indeed a channel refinement.
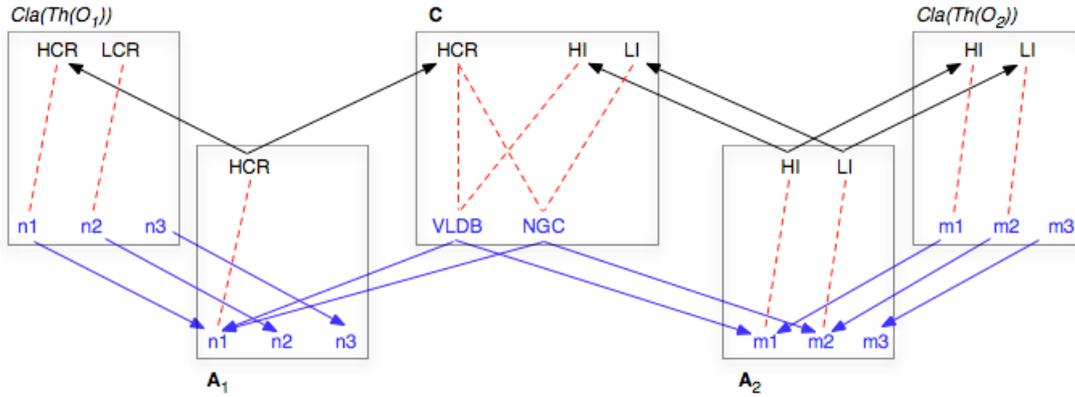


*Figure 2.* New partial alignment $\mathcal{A}$ in the Research Quality Assessment
scenario after exchanging new units of meaning coordination.

*Example (Research Quality Assessment).* Take the alignment shown in Figure 1 and assume that the agents exchange the units of meaning coordination "token 0288-3635 is of type high-citation-rate" and "token 0288-3635 is of type low-impact", using the ISSN of the journal *New Generation Computing*. Figure 2 shows the new alignment obtained (where we use the abbreviation NGC instead of the ISSN). The theory of the natural logic of the alignment's colimit will be the regular closure of the following set of constraints:

$$\begin{array}{lll}
 & \vdash & \text{high-citation-ratio}_1 \\
\text{low-citation-ratio}_1 & \vdash & \\
 & \vdash & \text{high-impact}_2 \,,\, \text{low-impact}_2 \\
\text{high-impact}_2,\, \text{low-impact}_2 & \vdash &
\end{array}$$

This theory models a more refined semantic integration than the one determined before the additional exchange of units of meaning coordination.


## 5. An Ontology-Alignment Interaction Model

By exchanging units of meaning coordination two agents progressively align their ontologies, and in the previous section we have modelled this fact by showing that the information channel describing the flow of information between ontologies is actually refined. Since we have been neutral with respect to what exactly tokens and types in units of meaning coordination are, the framework above applies to several different approaches to meaning coordination. We have already pointed to this fact in Section 2. For a practical application of our framework to ontology alignment in open, distributed environments, in this section we will show how our framework serves as a foundation for a general ontology-alignment interaction model. We shall first describe the process of meaning coordination from an operational perspective, and then provide an executable specification of such interaction model by using LCC (Robertson, 2004), the executable interaction-model specification language that is currently used as the core interaction-model language in the *OpenKnowledge* Specific Targeted Research Project (OpenKnowledge, 2006), sponsored by the European Commission under its 6th Framework Program.

The strategy that each agent may follow in selecting appropriate units of meaning coordination will obviously influence the quality of the alignment that one eventually gets. In the process of meaning coordination we describe next, agents $A_1$ and $A_2$ alternate in exchanging units of meaning coordination (hereafter, UMC) in order to explain each other the meaning of local and foreign types. This process gradually builds up an alignment and is based on the following coordination tactic: if an agent $A_i$ wants to known the meaning of a foreign type, it asks agent $A_j$ for a token of this type in order to classify this token according to its own ontology; reciprocally, $A_i$ may inform $A_j$ which type he has selected for this particular token. This dialogue may be described schematically as follows:

Agent $A_i$ wants to know the meaning of $O_j$-type $\alpha$:
1. $A_j$ selects a new token $a$ for $O_j$-type $\alpha$
2. $A_j$ sends $A_i$ the UMC "token $a$ is of type $\alpha$"
3. $A_i$ selects an $O_i$-type $\beta$ for token $a$
4. $A_i$ sends $A_j$ the UMC "token $a$ is of type $\beta$"

At this point, both $A_i$ and $A_j$ may update the alignment $\mathcal{A}$ as defined in the previous section because the dialogue above involves the exchange of two UMCs. Also, $A_j$ may take the new $O_i$-type $\beta$ as starting point for an analogous dialogue in order to find out the meaning of this foreign type:

Agent $A_j$ wants to know the meaning of $O_i$-type $\beta$:
1. $A_i$ selects a new token $b$ for $O_i$-type $\beta$
2. $A_i$ sends $A_j$ the UMC "token $b$ is of type $\beta$"
3. $A_j$ selects an $O_j$-type $\gamma$ for token $b$
4. $A_j$ sends $A_i$ the UMC "token $b$ is of type $\gamma$"

Again, at this point, both $A_i$ and $A_j$ may update the alignment $\mathcal{A}$ as defined in the previous section, because again the dialogue involves the exchange of two UMCs. Notice that this second dialogue is identical to the first one, only with the roles of agents $A_i$ and $A_j$ switched.

In the meaning coordination process described above we have been deliberately silent on how agents select tokens and types for the UMCs they need to exchange, and also at the stage at which the alignment process finishes (e.g., because some good-enough alignment has been achieved). In an open, distributed system the strategy followed by agents will surely depend on the local decision-making machinery. Instead, we want to focus on the shared interaction model that agents would need to follow to coordinate their ontologies, independently of their particular decision-making strategies. For this, we need to supply, in an executable language, the specification of the general process of agent coordination that may yield (subject to the agents' strategies) to an alignment of ontologies, with the roles undertaken by the agents during that process. In the remainder of this section we give a specification of the meaning coordination described above in one such language, namely LCC (Robertson, 2004) (see Figure 3 for a detailed definition of LCC's syntax).

$$
\begin{aligned}
Interaction\_Model \;\; &:= \;\; \{Clause, \ldots\} \\
Clause \;\; &:= \;\; Agent :: Dn \\
Agent \;\; &:= \;\; \mathbf{a}(Role, Id) \\
Dn \;\; &:= \;\; Agent \mid Message \mid Dn \text{ then } Dn \mid Dn \text{ or } Dn \mid Dn \text{ par } Dn \mid \mathbf{null} \leftarrow C \\
Message \;\; &:= \;\; M \Rightarrow Agent \mid M \Rightarrow Agent \leftarrow C \mid M \Leftarrow Agent \mid C \leftarrow M \Leftarrow Agent \\
C \;\; &:= \;\; Term \mid C \wedge C \mid C \vee C \\
Role \;\; &:= \;\; Term \\
M \;\; &:= \;\; Term
\end{aligned}
$$

Where **null** denotes an event that does not involve message passing; *Term* is a structured term (e.g., a Prolog term) and *Id* is either a variable or a unique identifier for an agent.

*Figure 3.* Syntax of LCC interaction models

Each of Clauses 1 to 3, below, defines the message-passing behaviour of a role in the interaction. Clause 1 defines the message-passing behaviour of an agent (identified by $A_i$) in the role of an *aligner* of ontologies. An agent in this role initiates a dialogue with another agent in the same role with the objective of building on top of an alignment $\mathcal{A}^n$ (which initially, for $n = 0$, may be empty) through the exchange of UMCs.

When in the role of an *aligner*, agent $A_i$ either may choose to wait for a message from an agent $A_j$ asking $A_i$ to explain the meaning of a type $\alpha$, and switching subsequently to the role of an *explainer* of $\alpha$ for agent $A_j$; or else it may choose to send a message to an agent $A_j$ asking $A_j$ to explain the meaning of a type $\alpha$, and switching subsequently to the role of an *inquirer* of $\alpha$ for agent $A_j$.

$\mathbf{a}(aligner(\mathcal{A}^n), A_i) ::$
$\quad$( ask(explain($\alpha$)) $\Leftarrow \mathbf{a}(aligner(\_), A_j)$ **then**
$\qquad \mathbf{a}(explainer(\mathcal{A}^n, \alpha, A_j), A_i)$ )
$\quad$**or** $\hfill (1)$
$\quad$( ask(explain($\alpha$)) $\Rightarrow \mathbf{a}(aligner(\_), A_j)$ **then**
$\qquad \mathbf{a}(inquirer(\mathcal{A}^n, \alpha, A_j), A_i)$ )

Clause 2 defines the message-passing behaviour of an agent (identified by $A_i$) in the role of an *explainer* of a local type $\alpha$ for agent $A_j$. An agent in this role exchanges with its dual agent (the agent that switched to the *inquirer* role when both initiated the dialogue in the *aligner* role) a pair of UMCs in order to update its alignment $\mathcal{A}^n$ with the ontology of $A_j$.

When in the role of an *explainer* of a local type $\alpha$ for agent $A_j$, an agent $A_i$ first sends a message to agent $A_j$ (in the role of an *inquirer*) telling it that $a$ is a token of $\alpha$, conditioned to $A_i$ being capable of selecting such token $a$ for $\alpha$. Next, it sends a new message to $A_j$ asking it to classify $a$ according to $A_j$'s ontology. Then it waits for a message from $A_j$ telling $A_i$ that $a$ is a token of some foreign type $\beta$. $A_i$ then updates its current alignment $\mathcal{A}^n$ according to the exchanged UMCs (that $a$ is of type $\alpha$ and of type $\beta$), which yields the new alignment $\mathcal{A}^{n+1}$. Finally, $A_i$ may choose to either continue the alignment with $A_j$, switching to the role of an inquirer of foreign concept $\beta$ for agent $A_j$, or else it may choose to exit the dialogue by switching back to the initial *aligner* role with the updated alignment.

$$
\begin{aligned}
&\mathbf{a}(\text{explainer}(\mathcal{A}^n, \alpha, A_j), A_i) :: \\
&\quad \text{tell}(\text{is\_of\_type}(a, \alpha)) \Rightarrow \mathbf{a}(\text{inquirer}(\_, \_, \_), A_j) \leftarrow \text{select\_token}(\alpha, a) \quad \mathbf{then} \\
&\quad \text{ask}(\text{classify}(a)) \Rightarrow \mathbf{a}(\text{inquirer}(\_, \_, \_), A_j) \quad \mathbf{then} \\
&\quad \text{tell}(\text{is\_of\_type}(a, \beta)) \Leftarrow \mathbf{a}(\text{inquirer}(\_, \_, \_), A_j) \quad \mathbf{then} \\
&\quad \mathbf{null} \leftarrow \text{update}(a, \alpha, \beta, \mathcal{A}^n, \mathcal{A}^{n+1}) \quad \mathbf{then} \\
&\quad ( \mathbf{a}(\text{inquirer}(\mathcal{A}^{n+1}, \beta, A_j), A_i) \quad \mathbf{or} \quad \mathbf{a}(\text{aligner}(\mathcal{A}^{n+1}), A_i) )
\end{aligned}
\tag{2}
$$

That is, given two agents in the aligner role of an interaction, when one asks the other for an explanation of a type, the former will switch into the role of an inquirer (the one sending out the message asking for the explanation), while the latter will switch into the role of an explainer (the one getting the message). Explainer and inquirer agents then enter a dialogue in which they subsequently exchange UMCs, until they decide (according to their local decision-making machinery) to exit the dialogue, and fall back to the initial aligner role. While in the explainer or inquirer role an agent will only pass messages with its dual agent.

Clause 3 defines the message-passing behaviour of an agent (identified by $A_i$) in the role of an *inquirer* of a foreign type $\beta$ for agent $A_j$. An agent in this role exchanges with its dual agent (the agent that switched to the *explainer* role when both initiated the dialogue in the *aligner* role) a pair of UMCs in order to update its alignment $\mathcal{A}^n$ with the ontology of $A_j$.

When in the role of an *inquirer* of a foreign type $\beta$ for agent $A_j$, an agent $A_i$ first waits for a message of agent $A_j$ (in the role of an *explainer*) telling it that $b$, for instance, is a token of $\beta$, and subsequently waits again for a new message from $A_j$ that asks $A_i$ to classify $b$ according to $A_i$'s ontology. It then sends a message to $A_j$ telling it that $b$ is a token of local type $\alpha$, conditioned to $A_i$ being capable of selecting such type $\alpha$ for which $b$ is a token. Next, $A_i$ updates its current alignment $\mathcal{A}^n$ according to the exchanged UMCs that $b$ is of type $\alpha$ and of type $\beta$, which yields the new alignment $\mathcal{A}^{n+1}$. Finally, $A_i$ may choose to either continue the alignment with $A_j$, switching to the role of an explainer of local type $\alpha$ for agent $A_j$, or else it may chose to exit the dialogue by switching back to the initial *aligner* role with the updated alignment.

$$
\begin{aligned}
&\mathbf{a}(\text{inquirer}(\mathcal{A}^n, \beta, A_j), A_i) :: \\
&\quad \text{tell}(\text{is\_of\_type}(b, \beta)) \Leftarrow \mathbf{a}(\text{explainer}(\_, \_, \_), A_j) \quad \mathbf{then} \\
&\quad \text{ask}(\text{classify}(b)) \Leftarrow \mathbf{a}(\text{explainer}(\_, \_, \_), A_j) \quad \mathbf{then} \\
&\quad \text{tell}(\text{is\_of\_type}(b, \alpha)) \Rightarrow \mathbf{a}(\text{explainer}(\_, \_, \_), A_j) \leftarrow \text{select\_type}(b, \alpha) \quad \mathbf{then} \\
&\quad \mathbf{null} \leftarrow \text{update}(b, \alpha, \beta, \mathcal{A}^n, \mathcal{A}^{n+1}) \quad \mathbf{then} \\
&\quad ( \mathbf{a}(\text{explainer}(\mathcal{A}^{n+1}, \alpha, A_j), A_i) \quad \mathbf{or} \quad \mathbf{a}(\text{aligner}(\mathcal{A}^{n+1}), A_i) )
\end{aligned}
\tag{3}
$$

The three clauses above specify an executable interaction-model by which two agents align their ontologies by exchanging UMCs. Recall that the formal framework on which the interaction model is based is neutral with respect to what tokens and types are, and how each agent does the classification. Consequently, this ontology-alignment interaction model specifies a general protocol that can be instantiated to particular ontology-alignment scenarios as those discussed in Section 2. Being independent of the classifying and decision-making machinery each agent might have, it offers a general model of ontology-alignment, which different agents can subscribe to.

## 6. Discussion

In this article we presented a formal foundation of semantic alignment based on agent interaction. We opted for an agent-oriented approach, in which the capability of progressive ontology alignment is a need in distributed environments. The emphasis of this article was on the foundations for supporting general ontology-alignment interaction models. We argued that ontology alignment shouldn't be seen only from a functional perspective but as a basis for a continuous interaction between agents in dynamic environments. This sort of approach could supplement the majority of works in this area where ontology alignment is practiced as a one off task where agents have to align the entire input ontologies (or fragments of them) as requested.

By providing a sound theoretical ground upon which we base our three hypotheses for achieving semantic interoperability, we enable the use of our framework to model semantic-alignment as it occurs in semantic heterogeneity scenarios by applying a variety of technologies. Instead of exploring concrete instantiations of the formal model to particular alignment technologies—wandering into the discussion of particular choice methods, termination criteria and alignment algorithms—we decided to shift our attention to what basic capability an agent should have to be able to engage in an ontology-alignment interaction. Choice of tokens and types, interaction termination criteria, and concrete matching algorithms will play a central role when grounding the formal model in concrete domains.

Another direction for our future work is that of using the executable specification language LCC to experiment with different ontology alignment systems. The generic nature of LCC makes it possible to accommodate a variety of agent-based interactions, which in turn could, and should, be attached to the front-end of ontology alignment systems on the Semantic Web. Our work in this direction is still in its initial stages, but the formal framework and the foundations for reasoning we provide in this article is ready and already applied in various scenarios.

## References

Alagic, S., & Bernstein, P. (2002). A model theory for generic schema management. In G. Ghelli & G. Grahne (Eds.), Database Programming Languages, Lecture  Notes in Computer Science (Vol. 2397, pp. 228–246). Springer.

Bailin, S., & Truszkowski, W. (2002). Ontology negotiation between intelligent information agents. The Knowledge Engineering Review, 17(1), 7–19.

Barr, M. (1996). The Chu construction. Theory and Applications of Categories, 2(2), 17–35.

Barwise, J., & Seligman, J. (1997). Information Flow: The Logic of Distributed Systems. Cambridge University Press.

Bench-Capon, T., Malcolm, G., & Shave, M. (2003). Semantics for interoperability: relating ontologies and schemata. In V. Marík, W. Retschitzegger, & O. Stepánková (Eds.), Database and Expert Systems Applications, Lecture Notes in Computer Science (Vol. 2736, pp. 703–712). Springer.

Bouquet, P., Serafini, L., & Zanobini, S. (2003). Semantic coordination: a new approach and application. In D. Fensel, K. Sycara, & J. Mylopulos (Eds.), The Semantic Web — ISWC 2003, Lecture Notes in Computer Science (Vol. 2870, pp. 130–145). Springer.

CiteSeer (2003). Estimated impact of publication venues in computer science. Retrieved June 30, 2006 from http://citeseer.ist.psu.edu/impact.html

Corrêa da Silva, F., & Agustí, J. (2003). Knowledge Coordination. Wiley.

Doan, A., & Halevy, A. Y. (2005). Semantic integration research in the database community: A brief survey. AI Magazine, 26(1), 83-94.

Dunn, J. M., & Hardegree, G. M. (2001). Algebraic Methods in Philosophical Logic. Oxford University Press.

Enderton, H. (2002). A Mathematical Introduction to Logic (2nd ed.). Academic Press.

Farquhar, A., Fikes, R., & Rice, J. (1997). The Ontolingua Server: a tool for collaborative ontology

construction. International Journal of Human-Computer Studies, 46(6), 707–727.

Ganter, B., & Wille, R. (1999). Formal Concept Analysis. Springer.

Garfield, E. (1994). The ISI impact factor. Retrieved June 30, 2006, from http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/

Giunchiglia, F., & Shvaiko, P. (2003). Semantic matching. The Knowledge Engineering Review, 18(3), 265–280.

Goguen, J. (in press). Information integration in institutions. To be published in a memorial volume for Jon Barwise edited by L. Moss. Available from http://www.cs.ucsd.edu/users/goguen/

Goguen, J., & Burstall, R. (1992). Institutions: Abstract model theory for specification and programming. Journal of the ACM, 39(1), 95–146.

Journal Citation Reports (2005). Science Edition. Thompson Scientific.

Kalfoglou, Y., & Schorlemmer, M. (2003a). IF-Map: An ontology-mapping method based on information-flow theory. In S. Spaccapietra, S. March, & K. Aberer (Eds.), Journal on Data Semantics I, Lecture Notes in Computer Science (Vol. 2800, pp. 98–127). Springer.

Kalfoglou, Y., & Schorlemmer, M. (2003b). Ontology mapping: The sate of the art. The Knowledge Engineering Review, 18(1), 1–31.

Kalfoglou, Y., & Schorlemmer, M. (2004). Formal support for representing and automating semantic interoperability. In C. Bussler, J. Davies, D. Fensel, & R. Studer (Eds.), The Semantic Web: Research and Applications, Lecture Notes in Computer Science (Vol. 3053, pp. 45–60). Springer.

Kent, R. E. (2000). The information flow foundation for conceptual knowledge organization. In C. Beghtol, L. Howarth, & N. J. Williamson (Eds.), Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International ISKO Conference, Advances in Knowledge Organization (Vol. 7). Würzburg: Ergon.

Kent, R. E. (2005). Semantic integration in the information flow framework. In Y. Kalfoglou, M. Schorlemmer, A. Sheth, S. Staab, & M. Uschold (Eds.), Semantic Interoperability and Integration, Dagstuhl Seminar Proceedings (Vol. 04391). IBFI, Schloss Dagstuhl, Germany.

Lenat, D. (1995). CyC: A large-scale investment in knowledge infrastructure. Communications of the ACM, 38(11).

Noy, N. F. (2004). Semantic integration: a survey of ontology-based approaches. SIGMOD Record, 33(4), 65–69.

OpenKnowledge (2006). Retrieved June 30, 2006, from http://www.openk.org/

Pratt, V. (1995). The Stone gamut: A coordinatiztion of mathematics. In Proceedings, 10th Annual IEEE Symposium on Logic in Computer Science (pp. 444–454). IEEE Computer Society Press.

Pratt, V. (2001). Orthocurrence as both interaction and observation. In R. Rodriguez & F. Anger (Eds.), IJCAI-01 Workshop on Spatial and Temporal Reasoning.

Rahm, E., & Bernstein, P. A. (2001). A survey of approaches to automatic schema matching. The VLDB Journal, 10(4), 334–350.

Robertson, D. (2004). Multi-agent coordination as distributed logic programming. In B. Demoen & V. Lifschitz (Eds.), Logic Programming, Lecture Notes in Computer Science (Vol. 3132, pp. 416-430). Springer.

Schorlemmer, M., & Kalfoglou, Y. (2003). On semantic interoperability and the flow of information. In A. Doan, A. Halevy, & N. Noy (Eds.), Semantic Integration, Proceedings of the ISWC 2003 Workshop, CEUR Workship Proceedings (Vol. 82).

Schorlemmer, M., & Kalfoglou, Y. (2005). Progressive ontology alignment for meaning coordination: An information-theoretic foundation. In F. Dignum et al. (Eds.), Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (pp. 737–744). ACM Press.

Sheth, A., & Larson, J. (1990). Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM Computing Surveys, 22(3), 183–236.

Shvaiko, P., & Euzenat, J. (2005). A survey of schema-based matching approaches. In S. Spaccapietra, S. March, & K. Aberer (Eds.), Journal on Data Semantics IV, Lecture Notes in Computer Science (Vol. 3730, pp. 146–171). Springer.

Standard Upper Ontology Working Group (2003, December 18). Retrieved June 30, 2006, from http://suo.ieee.org/

van Diggelen, J., Beun, R.-J., Dignum, F., van Eijk, R. M., & Meyer, J.-J. (2006). ANEMONE: An effective minimal ontology negotiation environment. In P. Stone & G. Weiss (Eds.), Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (pp. 899–906). ACM Press.

Wang, J., & Gasser, L. (2002). Mutual online ontology alignment. In S. Cranefiel, T. Finnin, & S. Willmott (Eds.), OAS'02: Ontologies in Agent Systems, Proceedings of the AAMAS 2002

Workshop, CEUR Workshop Proceedings (Vol. 66).

Zimmermann, A., Krötzsch, M., Euzenat, J., & Hitzler, P. (2006). Formalizing ontology alignment and its operations with category theory. In B. Bennett & C. Fellbaum (Eds.), Formal Ontology in Information Systems, Proceedings of the Fourth International Conference (FOIS 2006), Frontiers in Artificial Intelligence and Applications (Vol. 150). IOS Press.