Building Semantic Channels between Heterogeneous Agents with Topological Tools

E.Valencia*, J-P Sansonnet**
* Oosteinde 173, 2611 VD Delft, Nederland - eroika@zonnet.nl
** LIMSI-CNRS, BP 133 F-91403 cedex - jps@limsi.fr

Abstract

Semantic Heterogeneity arises between independently designed and evolving informational agents. Homogeneity between the knowledge bases of the two interacting agents cannot be reached by simple fusion, since the nature of the interaction restricts information exchange to streaming. For each pair of interacting agents, we define a semantical channel as an asymmetrical structure, build incrementally by each agent. The semantical channel is a subset of the agent's knowledge base and corresponds to the knowledge shared with the other interacting agent. We provide a formal definition of semantic heterogeneity and of semantic channels, as well as an algorithm for incrementally building the semantical channels between any pair of heterogeneous agents using tools from algebraic topology.

Keywords: Autonomous Agents, Knowledge Representation, Ontologies.

1 Introduction

Agents technology [luck2003] constitute an increasingly performing tool for modern application domains such as automatization of tasks, management of information exchange, semantic web, etc.

These technologies rely on the Agent Model. An *agent* is an autonomous flexible computer system capable of action in an unpredictable, dynamic and open environment. It is an abstraction tool for the design and construction of the above systems. The crucial aspect about the agents abstraction is the study of the ways in which the agents may *interact*.

One of the identified challenges that agents technology now faces, is to provide a *semantic* infrastructure for open agent communities. The creation of common ontologies, or knowledge bases playing a central role. *Information* agents typically have access to multiple, *heterogeneous* and geographically distributed information sources, in the Internet and corporate intranets, and search for relevant information. Agents designed by different designers, for different purposes and at different moments in time, are by definition *semantically heterogeneous* agents. Fusioning the heterogeneous data of the interacting agents is not always possible, nor wishable, because the nature of the interaction imposes that information is exchanged by streaming and because the agent may belong more permanently to systems that are inheritently heterogeneous.

We propose that for each interaction, the pair of agents open two asymmetrical channels that model their point of view on this interaction with the other agent. The first and second channels are the perspectives of each agent on the interaction and are not equal. However, it is necessary that a common grounding ontology is reached, or attempted to be reached during this interaction, as shown in [valencia2000], to allow further exchange of information. The construction of such a common ontology is done incrementally *via* the semantic channels, which inturn are a reflection of this ontology.

2 Formalizing the Problem

2.1 Context, Contraints and Assumptions

The system is restricted to two semantically heterogeneous information agents and we naturally study the interaction at the information level. The information agents are characterized by their *knowledge* bases constituted by an *ontology* base and an *assertions* base, by analogy with the Tbox/Abox scheme well known in Description Logics [donini1995]. The ontology base is the abstraction structure reflecting the resulting point of view of an agent on a world W at a time t. The assertion base is its expression of this world in terms of concepts of the ontology. The construction of the one depends on the construction of the other.

The two information agents interact in W and form part of a common system. The implementation of this wider system is a classical multi-agent platform not presented in this paper (see [sansonnet2003] for n implementation of a Multi-agent system framework dedicated to informational agents).

The portion of the world that is available to the perception of the agents is agent-dependent, and the assertions that result from their perceptions do not necessarily describe the same portions of the world. However, we maintain that there is in principle a minimum level of consistency that is guarantied by the fact that the knowledge bases of the two agents are in a way "provoqued" by a consistent source, W. The assertion bases of the agents use their ontology which are semantically heterogeneous by assumption.

As a consequence, we postulate a different common sub-ontology for each agent, as containing the concepts this agent *deduced* as being shared, according to deduction criteria that we will discuss in section 3.2. These sub-ontologies are called the *shared base* of an agent. The shared base is not defined as a substructure of the knowledge base of an agent, but is a structure attached to the channel that the agent has opened with the second agent.

The goal of a simulation is to construct a common *sub-ontology*, with concepts and relations between concepts that are tested to be shared, only by using the interaction process itself.

2.2 Notations

In this paper, we note A_1 and A_2 the interacting agents. When talking about generalization, any other agent will be A_i.

An interaction is noted $I(A_1; A_2; W; T)$ where W is a *world* of objects in which the agents interact, and T is a duration in units of interaction steps. The world$W$ is a set of *individuals* noted $\#j$, where $\#$ is meant to indicate an absolute reference, not depending on the naming of these individuals in the local knowledge bases of the agents, and j is an integer identifier.

An interaction is a sequence of interaction steps, noted i_t, where t is the step number $t \in [0, T]$. The interaction steps are simply requests and responses to the requests.

The agents possess a knowledge base, noted Kb_1 for A_1 and Kb_2 for A_2. The two components of a knowledge base Kb_1 are noted Tb_1 and Ab_1 by analogy with description logics because Tb_1 defines the ontological base and Ab_1 defines the assertions about W using concepts of Tb_1.

The Tb_1 base is a list of definition of concepts together with and their structural links.

The Ab_1 base is a list of assertions about the perceived parts of the world W. It comprises facts, relations and rules using concepts of Tb_1. The base Ab_1 contains facts f_1 and $n-$ary relations r_1.

In general the subscript i in X_i relates to the agent A_i in characterizing the structure X, while the superscript X^t relates to the time step t.

We will extensively use the connectors of the classical first order logic to introduce definitions in the knowledge bases.

When we refer to the simplicial representations of the structures defined above, we always use the suffix σ in front of the name of the structure. For example, the simplicial representation of Tb_1 is noted σTb_1.

2.3 Semantic Heterogeneity

It is not the **origin** of semantic heterogeneity between two information agents that is discussed here [minsky1985], but rather its consequences. More information about the "Weak Alien" Project at LIMSI-CNRS can be found on the web page: [sansonnet2004]. When we postulate a heterogeneity between the Kb_i, it means a heterogeneity in the semantics of the ontologies Tb_i. The semantics is here seen as the extension definition of a concept, in the sense of description logics, see [donini1995]. This heterogeneity at the ontology level will have consequences also at the assertions level, and the bases Ab_i are also heterogeneous in that sense. However, the semantics of each concept is only defined in Tb_i and *used* in Ab_i.

2.4 Semantic Channel

Classically [weinstein1999], an interaction I is characterized by the interacting agents A_i, the world W in which they interact, and the duration t in units of steps, of the interaction. An instance of interaction is then a specification of $(A_1; A_2; W; t)$.

We propose here that the interaction is further *reified*: not only the elements that allow the interaction to occur are characterizing the instance of interaction, but also the elements resulting from each step of this interaction. The structure reflecting this reification of an interaction is called *channel* by analogy with communication channels.

One channel is defined from A_1 to A_2, it is called C_{12}, and another channel is defined from A_2 to A_1, it is called C_{21}. Indeed, the channels have a direction and they characterize the interaction froim the point of view of one agent only. The channels C_{12} and C_{21} are not equal, there is not necessarily a symmetry. One *channel* is attached to each agent for each new interaction. This channel comprises three layers called physical, historical and semantical channels, we note $C_{12} = \{pc_{12}; hc_{12}; sc_{12}\}$, where:

1. Physical Channel pc_{12}: this channel corresponds to the hardware and software communication channels that support the interaction, it is not discussed here;

2. Historical Channel hc_{12}: this channel is the record of the history of the interaction, attached to each agent. This history records what information has already been send, what information has been agreed, rejected, or was already known by the receiving agent. In [sansonnet2003], a value of trust is also calculated for the receiving agent depending on the exchange of information at a time t.

3. Semantical Channel sc_{12}: this last layer expresses also *results* about the information exchanged but at the *semantic* level. This layer is the core of the proposition of this paper. The semantical channel is a composed structure, just like the historical channel above. It contains a pile of hypothesis h_1, a record of dialog acts send with respect to these hypothesis Act_1, and a shared base Sb_1 at each time t, $sc_{12} = \{h_1; Act_1; Sb_1\}$.

The historical layer plays a crucial role in the construction of the semantic layer because we wille see that the shared base Sb_1 is build using the hypothesis of h_1, and also the records about the information exchanged from the historical channel hc_{12}.

λ	a_1	a_2	a_3
b_1	1	0	0
b_2	0	1	1
b_3	1	1	1

Table 1: Incidence matrix M_λ of the binary relation λ.

2.5 Simplicial Representation

We propose a knowledge representation model of Ab_1 to Tb_1 in terms of simplicial complexes, that are a basic structure of algebraic topology. The corresponding simplicial representation model that we use here, was introduced by [valencia2000].(see also [lahiri2000] for a first course). We recall here some basic definitions of this model.

2.5.1 Definitions

Let $s = \{a_0, a_1, ..., a_p\}$ be a geometrically independent set in R^n.

Definition The *geometric p-simplex*[hatcher2002] generated by s, noted $\langle s_p \rangle$, is defined by $\langle s_p \rangle = \Sigma_{i=0}^{p} \lambda_i a_i$ with $\Sigma_{i=0}^{p} \lambda_i = 1$ and where $\lambda_i \in [0, 1]$ and $i = 0, 1, ..., p$.

A 0-simplex is a point ; a 1-simplex $\langle a_0, a_1 \rangle$ is an open interval joining a_0 to a_1 ; a 2-simplex $\langle a_0, a_1, a_2 \rangle$ is the inside of the triangle formed by $\{a_0, a_1, a_2\}$; etc.

For all simplexes $\langle s_p \rangle$ defined by $\langle a_0, a_1, ..., a_p \rangle$, the elements $\{a_0, a_1, ..., a_p\}$ are called the edges of $\langle s_p \rangle$ and p its dimension.

If $\{i_0, i_1, ..., i_k\}$ is a subset of $\{a_0, a_1, ..., a_p\}$ then the simplex $\langle s_k \rangle$ defined by $\langle a_{i_0}, ..., a_{i_k} \rangle$ is a k-face of $\langle s_p \rangle$, we note it $\langle s_k \rangle \leq \langle s_p \rangle$.

Definition A *simplicial complex*[hatcher2002] K of R^n is a finite collection of open geometrical simplexes in R^n satisfying the conditions C_1 et C_2 :

1. C_1 : If $\langle s_p \rangle$ is a simplex of K and $\langle s_k \rangle <= \langle s_p \rangle$ then $\langle s_k \rangle$ is in K.
2. C_2 : If $\langle s_p \rangle$ et $\langle s_q \rangle \in K$ and $\langle s_p \rangle \neq \langle s_q \rangle$ then $\langle s_p \rangle \cap \langle s_q \rangle = \emptyset$.

As a consequence of C_1, if $\langle s_p \rangle$ is in K then all the faces of $\langle s_p \rangle$ are also in K. The dimension of K is defined as the maximum dimension of its simplexes. The notion of complex generalizes the notion of graph, since all the complexes ofdimension $n < 2$ sont are graphs.

2.5.2 Representing Binary Relations

Incidence Matrix M_λ Let λ be a binary relation from $A = \{a_1, a_2, a_3\}$ in $B = \{b_1, b_2, b_3\}$, defined by its incidence matrix M_λ (see table 1).

Simplicial Complex K_λ Let us represent the elements a_1, a_2 and a_3, with the 0-simplexes $\langle a_1 \rangle$, $\langle a_2 \rangle$ and $\langle a_3 \rangle$, like in [atkin1977]. Each element of B can then be represented as a p-simplexe defined by the 0-simplexes representing the elements of A with which they are λ-related. The set of simplicial representations $\langle a_1 \rangle$, $\langle a_2 \rangle$, $\langle a_3 \rangle$, $\langle b_1 \rangle$, $\langle b_2 \rangle$, $\langle b_3 \rangle$, is noted K_λ,and σK_λ is shown on figure 1. We have :

$$K_\lambda = \langle \langle a_1 \rangle, \langle a_2 \rangle, \langle a_3 \rangle, \langle a_2, a_3 \rangle, \langle a_1, a_2, a_3 \rangle \rangle$$

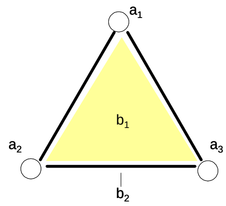

Figure 1: Simplicial representation of the binary relation λ.

λ	#1	#2	#3	#4	#5	#6	#7	#8	#9
odd	1	0	1	0	1	0	1	0	1
even	0	1	0	1	0	1	0	1	0
/3	0	0	1	0	0	1	0	0	1
/4	0	0	0	1	0	0	0	1	0

Table 2: Perception of $W = \{\#1, \#2, \#3, \#4, \#5, \#6, \#7, \#8, \#9\}$ by A_1 using the set of perceptors $P_1 = \{odd, even, /3, /4\}$.

Dual Representation: We could also start with taking the elements of B as 0-simplexes and representing the elements of A as p-simplexe based on them. The set of simplicial representations $\langle a'_1 \rangle$, $\langle a'_2 \rangle$, $\langle a'_3 \rangle$, $\langle b'_1 \rangle$, $\langle b'_2 \rangle$, $\langle b'_3 \rangle$, is noted K'_λ. We say that K_λ and K'_λ are duals, they represent the same relation and "show" the same information about it. This will be more apparent in the following examples.

2.5.3 Representing the Ontology Base

Concepts from W: An agent A_1 perceiving objects $\#n$ of the world W will abstract concepts concerning these objects. We propose that the abstractions that an agent is able to construct depend on its perceptors, and that its perceptors may be modelled by binary relations between the set of objects of W and the set of its perceptors, noted P_1.

Let for example take $P_1 = \{odd, even, /3, /4\}$ and $W = \{\#1, \#2, \#3, \#4 \#5, \#6, \#7, \#8, \#9\}$ where the individuals $\#i$ are the integers $1, 2, 3, , 5, 6, 7, 8, 9$. The *perception* of W by A_1 is the binary relationP_1 characterized by the incidence matrix M_{P_i}, given in table 2.

Each agent will constitute an abstraction of W by this process, and constitute its Tb_i, as proposed by [valencia2000].

Relations between concepts: The concepts are defined by the relation that they have with others in an ontology. Let us consider the main relation of *inheritance*. A concept C_2 that inherits from C_1 inherits all the *attributes* of C_1 and is *differentiated* from C_1 by at least one *attribute*.

Let us consider the concepts $Tb_i = \{Bird, Animal, Dog, Turtle\}$. Let us take the following *intensive* definition of these concepts:$bird = [wings]$, $dog = [tail, hair]$, $turtle = [tail]$[1]; and we also have the additional attributes *yellow*, *white*, *green*. Now we give an extensive definition of the concepts using the set of individuals $W = \{titi; coco; milou; paloma; carla\}$, see table 3, and the simplicial representation, see 2.

[1] According to this ontology, a dog is a kind of turtle!

Tb_i	*titi*	*coco*	*milou*	*paloma*	*carla*
yellow	1	0	0	0	0
green	0	1	0	0	1
white	0	0	1	1	0
wings	1	1	0	1	0
tail	0	0	1	0	1
hair	0	0	1	0	0

Table 3: Extensive definition of $Tb_i = \{Bird, Animal, Dog, Turtle\}$ with $W = \{titi; coco; milou; paloma; carla\}$.

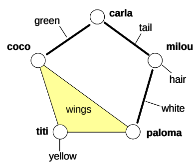

Figure 2: Simplicial representation of $Tb_i = \{Bird, Animal, Dog, Turtle\}$.

2.5.4 Representing assertions

We now assume that the two agents A_1 and A_2 have a knowledge base that is not empty, and that results from either a direct design by a designer prior to its entering into W, or its perception of other worlds before W. In the next subsection, we propose now to represent assertions base Ab_i using the simplicial complexes.

Representing facts Facts $f_1(\#n)$ of Ab_1 over elements $\#n$ of W are predicates, to represent facts $f_1(\#n)$ about $\#n$ is to build the incidence matrix between the set of predicates of the ontology Tb_i of the agent, and the set of elements of W. Since the agents are also interacting, and hopefully learning, at any time, we consider that the ontolgy Tb_i is not only the result of the application of the set P_i to W.

Representing $n-$ary relations n-ary relations $r_1(\#n_1, \#n_2, ..., \#n_n)$, with $n > 2$ are not so trivially representable, because to use Atkin's method[atkin1977], we need to start from a two dimensional incidence matrix. We propose to represent all the $n-$uples $(\#n_1, \#n_2, ..., \#n_n)$ of R^n as $0-$simplexes, and the n-ary relation is then seen as a predicate over the $n-$uples, that is, a binary relation from R^n in R^n, whose incidence matrix M_{r_1} is $2-$dimensional.

Representing rules A rule is of the form $L(\#n) \rightarrow R(\#n)$. Clearly, like with the *set* representation of rules, the operation of $\rightarrow$ is reflected by the *inclusion* of the representations of $L(\#n)$ and $R(\#n)$. In other words, if $\#n$ is a vertice of the complex representing the predicate L, noted σL, then $\#n$ will also be a vertice of the complex representing R, noted σR: $\sigma R \subset \sigma L$.

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
h_1	$\emptyset$	$p_2=a_1$ $p_2=b_1$	$p_2=a_1$ $p_2=b_1$	$\emptyset$	
Sb_1	$\emptyset$	$\emptyset$	$\sigma p_2=\sigma a_1$	$\sigma p_2=\sigma a_1$	$\sigma p_2=\sigma a_1$ $\sigma q_2=\sigma\#4$
heu	rd(Ab_1)	tst(h_1)	tst(h_1)	rd(Ab_1)	rd(Ab_1)
s_{12}	$b_1(\#1)$	$b_1(\#2)$	$a_1(\#3)$	$a_1(\#2)$	$a_1(\#1)$
h_2	$b_1=p_2$	$b_1=p_2$	$a_1=p_2$	$a_1=p_2$	
Sb_2	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\sigma a_1=\sigma p_2$
heu	tst(h_2)	tst(h_2)	rd(Ab_2)	rd(Ab_2)	rd(Ab_2)
s_{21}	$p_2(\#1)$	$p_2(\#2)$	$p_2(\#3)$	$q_2(\#4)$	$r_2(\#2)$

Table 4: Sequence of hypothesis h_1, action Act_1 and sentence s_{12} send by A_1, consequence on A_2 in terms of hypothesis h_2, action Act_2 and sentence s_{21} send to A_1 as a function of t.

3 Simulation

3.1 Basic Interaction Protocol

To simulate the interactions, we extend the multi-agent platform proposed by [sansonnet2003] with the Mathematica[wolfram1999] programming language.

Let A_1 send a sentence $s_{12}^{t=1}$ to the agent A_2, and let A_2 react with $s_{12}^{t=1}$. $s_{12}^{t=1}$ is called a *sentence* to differentiate it from a query, because s_{12} transports a piece of *content* information. The sentences are written in a very simple Communication Language, ACL[fritzson1994].

3.2 Incrementally building Semantic Channels

The semantic channel C_{12} is defined by:

$$C_{12} = \{h_1; Act_1; Sb_1\}$$

where Sb_1 is of the same nature than Tb_1, and is thus a list of definition of concepts together with and their structural links, its representation with simplicial complexes is noted σSb_1. We give an example of incremental construction of C_{12} and C_{21}. For clarity of the figures, we go back to abstract notations. Let:

$I = \{A_1, B_1, W, 5\}$,
$W = \{\#1, \#2, \#3, \#4\}$,
$Tb_1 = \{a_1, b_1\}$,
$Ab_1 = \{a_1(\#1), a_1(\#2), a_1(\#3), b_1(\#1), b_1(\#2)\}$,
$Tb_2 = \{p_2, q_2, r_2\}$ and
$Ab_2 = \{p_2(\#1), p_2(\#2), p_2(\#3), q_2(\#4), r_2(\#2)\}$.

Table 4 gives the sentences that are exchanged between A_1 and A_2 as a function of t, for the first 5 steps of an interaction. We explain synthetically below each step $t =$:

1. A_1 has an empty pile of hypothesis h_1, the function heu selects an information to be send, here it is calling rd which implements a *random selection*. The fact $b_1(\#1)$ is selected from Ab_1 and send as $s_21^{t=1}$ to A_2. At this same time step, the shared base of Sb_1 is still empty, $Sb_1 = \emptyset$. When A_2 receives $s_21^{t=1} = b_1(\#1)$, it makes an hypothesis about the heterogeneous terms in $s_21^{t=1}$,that is b_1.This is done as fommows:A_2 detects that b_1 is a fact about #1, ince A_2 has also a fact about #1, which is $p_2(\#1)$, it generates the

hypothesis that $b_1 = p_2$ which is then added to h_2. The heursitic `heu` used in turn by $A_2$to select an information to send is now noted `tst`, for *test the hypothesis.* This function selects a fact concerning p_2 because it is the part of the hypothesis that intersects with Ab_2, and the sentence $s_{21}^{t=1} = p_2(\#1)$ is send to A_1.

2. A_1 receives $s_{21}^{t=1}$ and generates the two hypothesis $p_2 = a_1$ and $p_2 = b_1$, and adds them to h_1[2]. The heursitic `tst` allows to select a fact about b_1 to send in $s_{12}^{t=2}$. A_2 receives $s_{12}^{t=2}$, generates no additional hypothesis, because the only deduction it can make from $s_{12}^{t=2}$ is already present in h_2. Again, the heuristic `tst` allows to select a fact about the hypothesis in h_2, and A_2 sends $s_{21}^{t=2} = p_2(\#2)$.

3. A_1 receives $s_{21}^{t=2}$ and generates no new hypothesis because the only deductions it can make from $s_{21}^{t=2}$ are already present in h_1. The heursitic `tst` allows to select a fact about the hypothesis in h_1, no more facts about b_1 can be sent since an agent cannot send the same fact more than once to the same agent. A fact concerning the other hypothesis is selected, i.e. about a_1. Then $s_{12}^{t=3} = a_1(\#3)$. A_2 receives $s_{12}^{t=3}$, and generates the new hypothesis $a_1 = p_2$. In order to test h_2, another fact using p_2 is selected and $s_{21} = p_2(\#3)$ is send to A_1.

4. A_1 receives $s_{21}^{t=3}$ and deducts that the hypothesis $p_2 = b_1$ cannot hold, then the hypothesis $p_2 = a_1$ is taken as valid, and added to Sb_1. The heuristic `rd` allows to send a new sentence since the hypothesis pile is now empty. Note that the hypothesis $p_2 = a_1$ is taken to be valid, when it has not strictly been proven. In fact it is not possible to prove a hypothesis in the context of interacting agents, we only deal with credible assumptions. Here, since two candidates were possible to be identified with p_2, as soon as one of them is eliminated, the other is confirmed. There are other choices possible, like setting a threshold of 3 confirming facts to allow a hypothesis to enter Sb_1.

5. When A_1 receives the fact $q_2(\#4)$ from A_2 there is nothing in Ab_1 that can match q_2 since A_1 possesses no facts about #4. It is therefore checked whether the *new* concept q_2 is *consistent* with Ab_1, if yes, the fact is *agreed*[3]. This new concept is directly added to Sb_1.

3.3 Simplicial representation

To illustrate the construction of Sb_1 and Sb_2 we show the simplicial representations σSb_1 and σSb_2 and σAb_1 and σAb_2, for comparison. We see that Sb_1 and Sb_2 are not identical, and have not been generated at the same interaction step. We can see from table 4 that they evolve. However, they seem to converge.

3.4 Algorithm

The general algorithm that takes $\{Tb_1; Ab_1; C_1; W, T\}$ as input and gives $\{C_{12}\}$ as an output is defined in the following main steps:

Inititialization, $t = 0$: `init`(h_1) function: creates empty data structures for h_1, Sb_1 and Act_1, initializes the records in hc_1 of C_{12} (see section 2.4 above).

[2] Two hypothesis of the same priority are tested in random order.
[3] This dialog act form part of the ACL of the platform that has not been presented here.

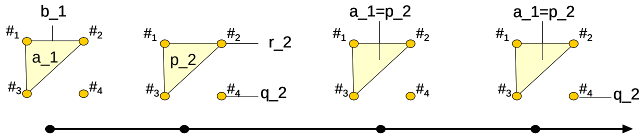

Figure 3: Simplicial representation of the two assertion boxes Ab_1 and Ab_2, and the two shared bases Sb_2 and Sb_1,after 5 interaction steps. Sb_2 and Sb_1 are not equal.

loop over $t > 0, t < T$:

1. `rec`(s_{21}) function: initialize list hs_t of heterogeneous labels at step t; syntaxic parsing of the content of s_{21}, with grammar Tb_1; if item not recognized, added to list of heterogeneous terms hs_t;

2. `hs`(hs_t) function: checks if individuals of the facts or relations in hs_t have facts or resp. rules attached in Ab_1; if none, call `deduct`(hs_t,Ab_1); if yes call `hyp`(hs_t, h_1).

3. `deduct`(hs_t,Ab_1) function: checks logic consistency between base Tb_1 and hs_t; if absurd triggers `reject` dialog act, if not pass hs_t to Sb_1;

4. `hyp`(h_1) function: checks the hypothesis pile $h_1 = \{h_1(i), ...\}$ of C_1; if empty then calls the `rd`(Ab_1) heursitic; if h_1 not empty, if items $h_1(i)$ with priority `prio`$(h_1(i)) = 3$ then pass item $(h_1(i))$ to Sb_1; if not take highest priority item $h_1(j)$ with `prio`$(h_1(j)) = max$ and calls the `tst`(Ab_1,$h_1(j)$).

5. `tst`(Ab_1,$h_1(top)$) function: calculates candidate *cand* as syntaxic intersection $h_1(top) \cap Ab_1 = cand$; if $cand = \emptyset$ then increment priority `prio`$(h_1(top))$`++` in h_1; if not calls `send`(*cand*).

6. `send`(*cand*) function: triggers dialog acts in accordance with output from `tst`; in the example of this paper, the only dialog act is an `Inform`.

4 Discussion and perspectives

4.0.1 Comparison with the Semantic Web approach

Work undertaken by J Hendler *and al* in the DAML project and on the Web services within the DAMLS project [DAMLS2004], are based on the general idea of a "great unified world ontology" of services, that is supposed to be shared by everyone. In the wake of this work much propositions have been made in the Knowledge Engineering field where ontologies are considered in terms of DTDs, which means that the semantic heterogeneity is solved by a consensus supported by committees of the kind of the W3C etc. [W3C2004]. Of course, these approaches show true practical common sense [lakshmanan2003], [klein2003], but on the scientific level, they are not entirely satisfying because they do not take into account the dynamicity and openness underlying any distributed system like the Internet, where every day, in an independent and not coordinated way, many innovative services, are launched. In contrast to the main stream approaches; we propose to handle semantic heterogeneity between pairs of agents, by building locally shared sub-ontologies and not by fusioning knowledge bases in a unified world onlogy.

4.0.2 Use of topological tools

In this paper, we proposed to handle semantic heterogeneity with techniques based on algebraic topology. In our view, these tools should not be frightning, and are in fact attractive for at least two reasons: Firstly, simplicial complexes can be seen as a generalization of graphs and are extensively used in symbolic data processing. Note that there is a relation between SC-matching algorithms and graph-matching algorithms. Secondly, simplicial complexes present a suitable visual representation of paths and borders between concepts, even though they are *in fine* reducible to boolean tables (see table 1).

Moreover, we explored the transposition of XML documents in Description Logics (DL) [borgida1996], [donini1995] and then in our simplicial representation formalism[4], using the correspondinces between DL and CAT already defined in earlier work. This study lead us to propose a direct transposition XML $\leftrightarrows$ CS, to be published in future works. From this, we concluded that the results achieved using one of these formalisms are in fact always transposable into the others, and therefore, the question of the comparison of our approach is not so much relevant at the level of the formalism itself, but rather at the level of our basic postulate: « *there exists and there will always exist a certain amount of semantic heterogeneity between the knowledge bases of informational agents, when they are distributed in a dynamic and open world* ».

4.0.3 Hypothesis of direct reference and sequential communication

In this work, for sake of simplification, we made the assumption that objects can be referenced by agents with direct reference tha is with unique IDs. It is true that associative reference (that is were entities are refered to via a combination of attributes) is a more likely situation to happen in MAS (Multi-agent Systems); we discussed this issue of "associative extensional reference" in [sansonnet2003a] and additional information can be found at [sansonnet2004].

Also, a frequently asked question is: since the communication between two agents is supposed reliable, why not sending all the information one has about an object at once? The answer lies in the informational agents paradigm where we don't want agents to exchange all the content of their knowledge bases at least for computational reasons. Therefore, agents try to exchange only the amount of information needed for their goals, thus resulting in a sequence of interaction steps instead of blunt 'downloads'.

4.1 Perspectives

In the proposed algorithm, each hypothesis that has not been contradicted for more than 3 steps is passed to Sb_1. In a more elaborate version, we propose to have a pile h_1 with a more complex structure. Each hypothesis item is not simply a equation between facts or relations, but uses domain restrictions. For example, in Tb_1, we have $b_1 = a_1|\{\#1, \#2\}$ meaning that over $\{\#1, \#2\}$, we have $a_1 = b_1$. This restriction may be used to make a hypothesis about mapping between concepts of Tb_1 and Tb_2. For example, at step $t = 5$, $p_2|\{\#1, \#2\} = b_1$.

The representation of knowledge with simplicial formalism is on the one hand seducing by its simplicity, however, a unique representation of a Kb_i is not achievable in a simple way. In particular, many types of relations require different complexes to be expressed. The main force of algebraic topology is to be able to compare spaces according to their topological properties. Even if the simplicial representation of a knowledge base seems complex, the calculation of topological indices of the bases (see [hatcher2002]) may lead us to deduct interesting properties about the bases of two agents that have been generated by the same world.

[4] To achieve that, the AL language is sufficient.

This work is a preliminary proposition, and a lot of work need to be done at the theoritical level, mainly to make sound proofs of the proposed algorithms. In particular, we made many simplifying hypothesis w.r.t. the forms of heterogeneity considered. Mainly, we assumed a common domain of representation with a common usage of constant names between the two representations, and presented a methodology to discover equivalence or equivalence restricted to a sub-domain between unary predicates belonging to different representations. In further work we need to tackle these serious limitations, which confirm that the problematics of semantic heterogeneity is a long run process.

References

[atkin1977] R.H. Atkin. Combinatorial Connectivities in Social Systems. Verlag, 1977

[lahiri2000] B.K. Lahiri. A first course in algebraic topology. Alpha Science International Ltd, 2000.

[donini1995] D. Donini, M. Lenzerini. Principles of Knowledge Representation, chapter in Reasoning in Description Logics, p193-238. CSLI Publications, 1995.

[fritzson1994] R. Fritzson, T. Finin, D. Mc Kay, R. McEntire. KQML: A language protocol for knowledge and information exchange. 13th International Distributed Artificial Intelligence Workshop, July 28-30, Seattle,1994.

[shaerf1993] A. Schaerf. Query answering in concept-based knowledge representation systems : algorithms complexity, and semantic issues. PhD thesis, dipartamento di informatica e sistemistica, Universita di Roma, 1993.

[muninda2001] P. Muninda, M. Singh. Towards a formal theory of communication for multiagent systems. Proceedings of the International Joint conference on Artificial Intelligence, (IJCAI), 2001.

[torroni2002] P. Torroni. A study on the termination of negociation dialogues. Proceedongs of the First International Joint Conference on Autonomous Agents and Multi-Agents Systems, pages 1223–1230, July 2002 AAMAS'2002. Bologna, Italy.

[hatcher2002] A. Hatcher Algebraic Topology. Cambridge University Press, 2002.

[luck2003] C. Luck, P. McBurney. Agent Technology: Enabling Next Generation COmputing. AgentLink, 2003.

[weinstein1999] P. C. Weinstein and W. P. Brimingham. Comparing concepts in differentiated ontologies. KAW'99, 1999.

[valencia2000] E. Valencia. Algebraic Topology for Semantical Heterogeneity between Dialogical Agents. PhD Computer Science of University Parix-XI, Dec 2000.

[wolfram1999] S. Wolfram. The Mathematica book. Wolfram Media. Cambridge Universtity Press, 1999.

[hendler2000] J. Heflin, J. A. Hendler, Dynamic Ontologies on the Web, AAAI, pp 443-449, 2000

[DAMLS2004] DAML-S: Web Service Description for the SEMANTIC WEB, 2002.

[W3C2004] W3C, Semantic Web, http://www.w3.org/2001/sw/, 2004.

[lakshmanan2003] L. Lakshmanan, F. Sadri, XML Interoperability, In Proc. of the International Workshop on the Web and Databases (WebDB), San Diego, California, USA, 2003.

[borgida1996] A. Borgida, D. L. McGuiness. Asking queries about frames. KR'96, 1996.

[klein2003] M. Klein, N. Noy, A component-based framework for ontology evolution, Technical Report IR-504, Department of Computer Science, Vrije Universiteit Amsterdam, 2003.

[sansonnet2003] J.-P. Sansonnet, E. Valencia, Modelization and experimentation of conversational policies for non task oriented agents based on dissonance and relevance theories, EUMAS 03, The First European Workshop on Multi-Agent Systems, Oxford, dec 2003.

[sansonnet2003a] G. Pitel, J-P. Sansonnet, Towards a Functional Approach of the Extensional Resolution of Intrinsic Reference in Practical Dialogue Proc. of the Pacific Association for Computational LINGuistics 2003, Halifax Canada, aug. 2003.

[sansonnet2004] J-P. Sansonnet. The weak aliens home page : http ://www.limsi.fr/~jps LIMSI-CNRS, 2004.

[hewitt1986] C. Hewitt. Offices are open systems, ACM Transactions on Office Information Systems, 4(3):271–287, 1986.

[minsky1985] M. Minsky. Communication with Alien Intelligence, chapter Science and Alien Intelligence. Cambridge University Press, 1985.