

# **Weak Aliens Problem**

**Un modèle topologique pour la résolution de  
l'Hétérogénéité Sémantique entre Agents  
Informationnels**

**Jean-Paul Sansonnet**

LIMSI-CNRS, BP 133, F-91403 Orsay Cedex, France -- [jps@limsi.fr](mailto:jps@limsi.fr)

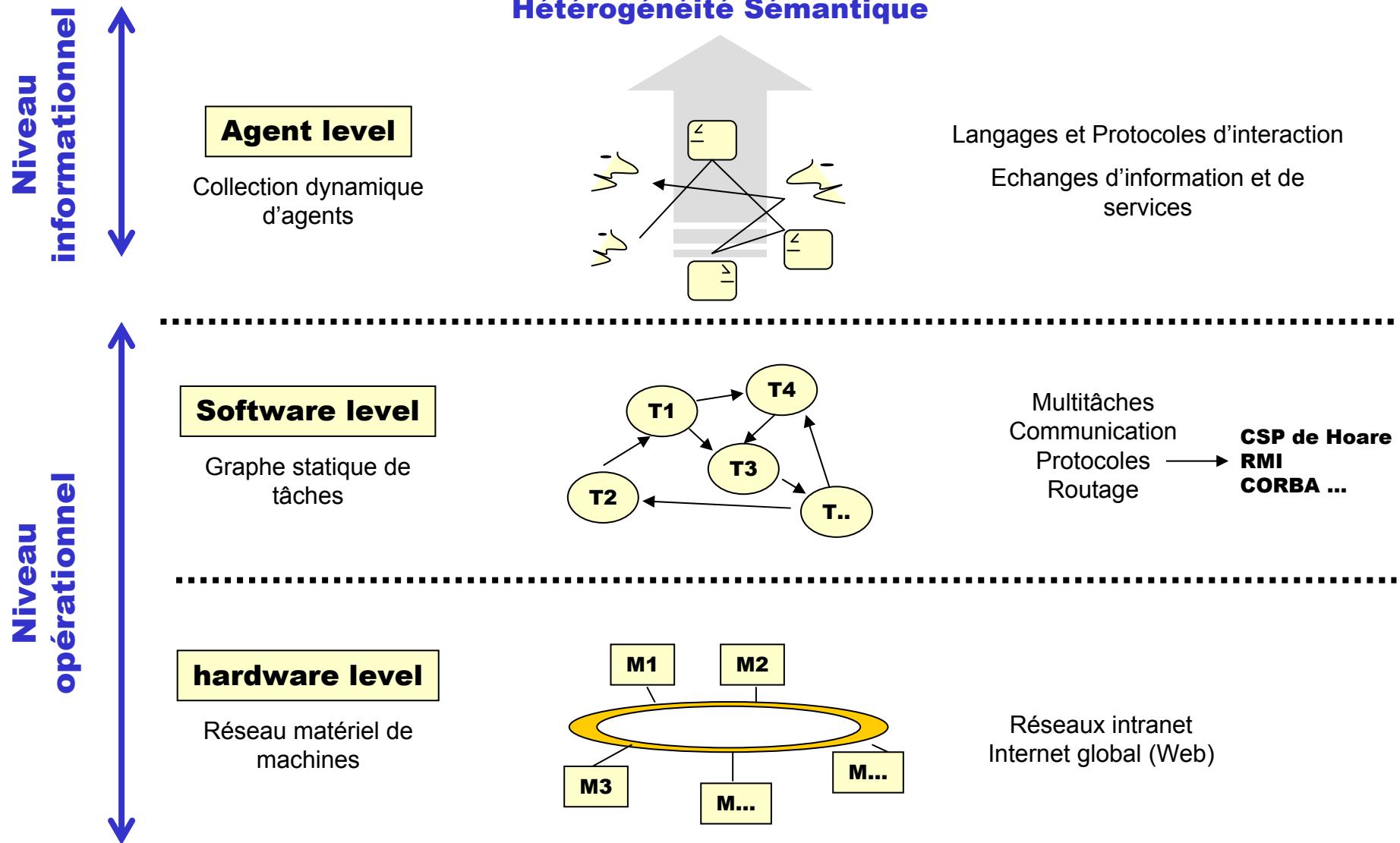
**Erika Valencia**

31, Wassenaarseweg 2596 CE Den Haag, PAYS-BAS -- [eroika@zonnet.nl](mailto:eroika@zonnet.nl)

# Les couches de la communication

2

## Problématique



# Les types d'hétérogénéité sémantique

3

## Problématique

Les sortes de différences entre ontologies ont été analysées par A. E. Campbell et S. C. Shapiro en 1995 :

### 1. Nommage

Cette hétérogénéité provient de l'utilisation d'un même nom (symbole) pour des objets différents ou bien des noms différents pour un même objet. On parle alors d'*hétérogénéité référentielle*.

### 2. Taxinomie

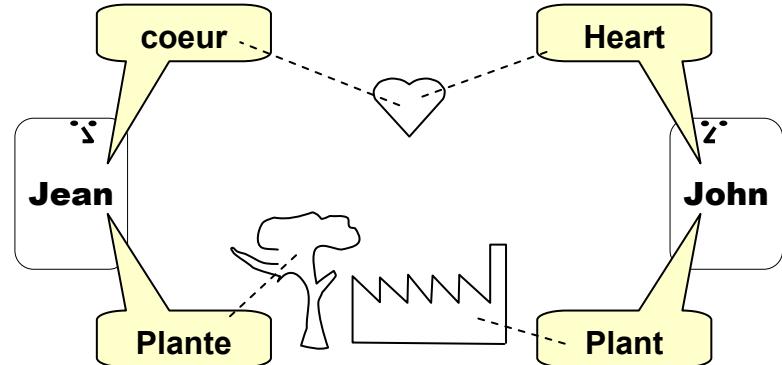
Cette hétérogénéité provient de la structuration en hiérarchie du domaine. On parle alors d'*hétérogénéité taxinomique*.

### 3. Structure

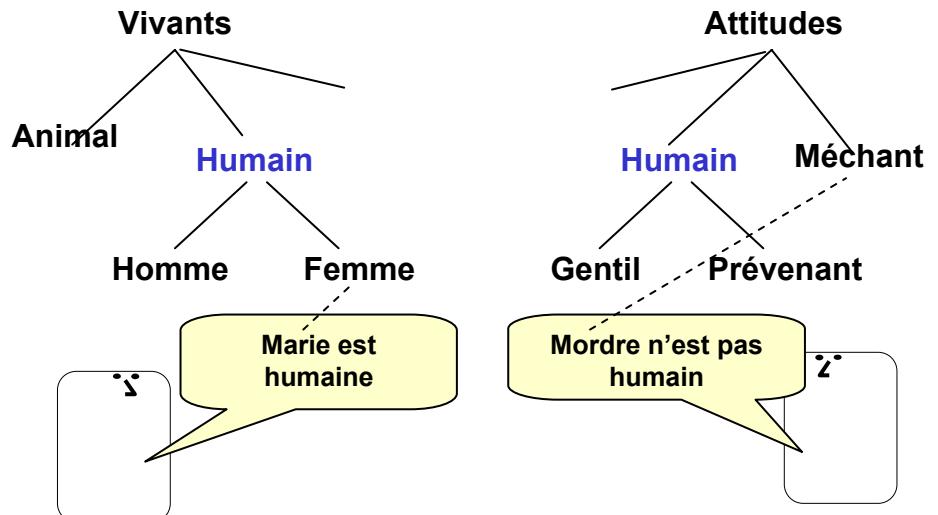
Cette hétérogénéité provient de la syntaxe employée pour transporter le contenu. Nous ne traitons pas ce type d'hétérogénéité ici car nous supposons le problème résolu via les infrastructures logicielles

### 4. Expressivité

Cette hétérogénéité provient du langage de représentation. Nous ne traitons pas ce type d'hétérogénéité ici car nous supposons le problème résolu via les infrastructures logicielles.



Hétérogénéité  
référentielle

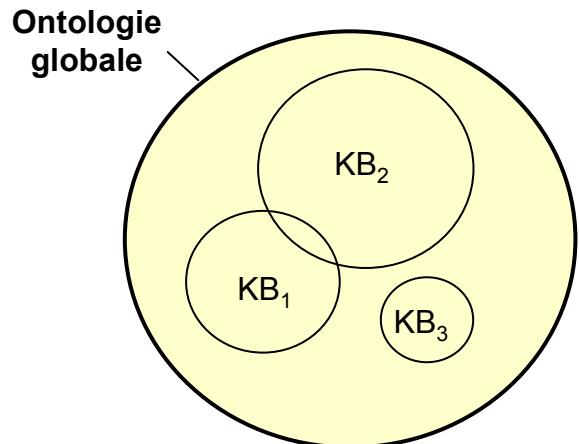


Hétérogénéité  
taxinomique

# Les trois situations de communication

4

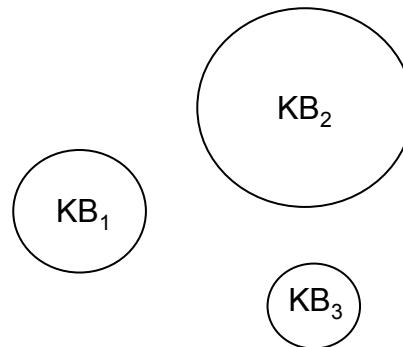
## Problématique



**NAP**

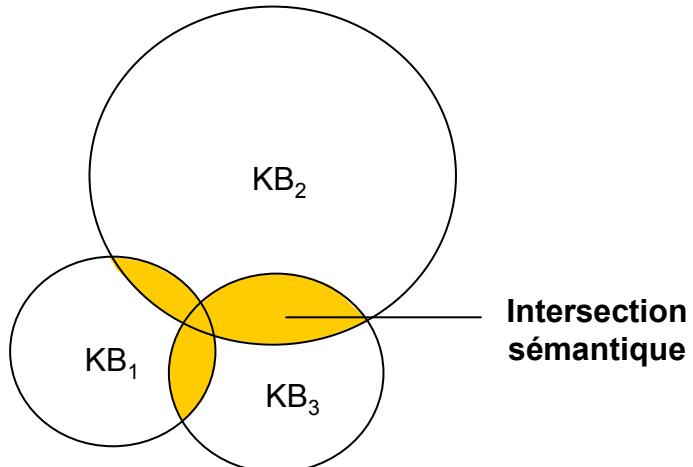
Ontologie globale qui subsume toutes les ontologies particulières et il n'a pas de problème de HS.

**SAP**



Les agents n'ont aucune intersection et ne peuvent absolument pas communiquer.

**WAP**

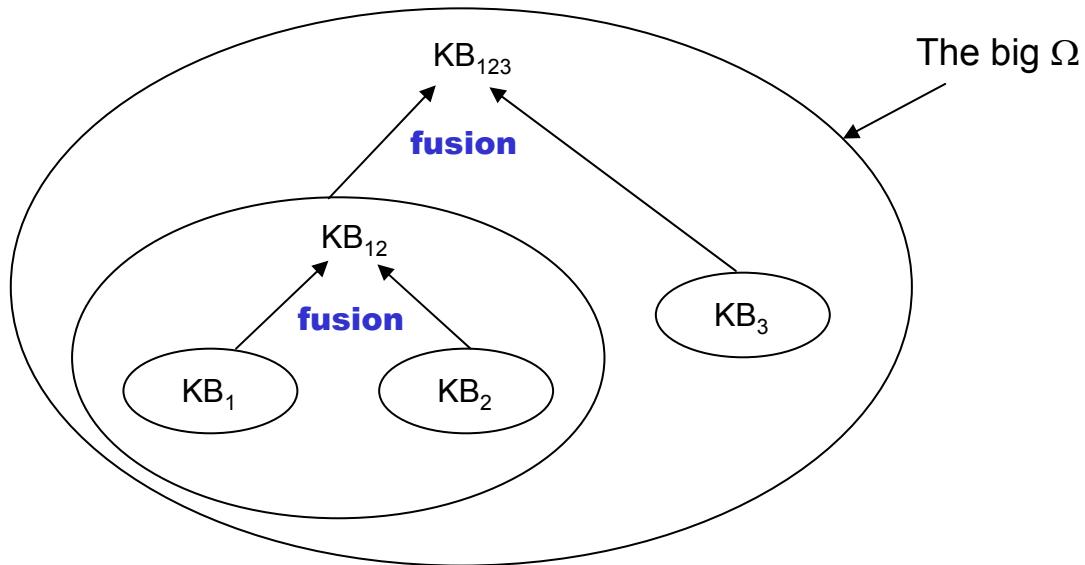


Les ontologies sont différentes mais ont cependant une intersection

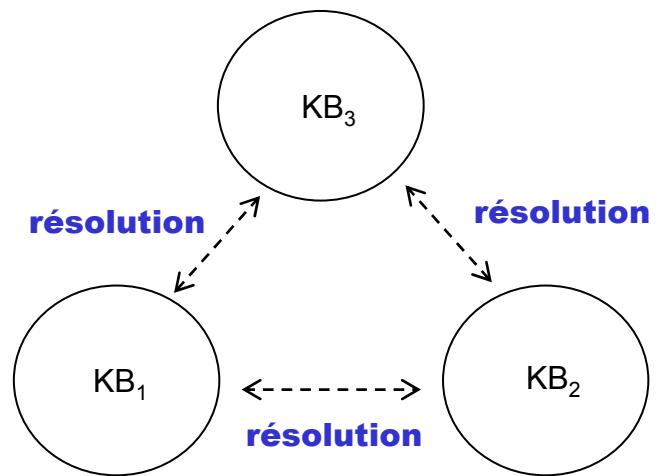
# Faut-il éradiquer ou affronter l'hétérogénéité ?

5

## Problématique



On peut **éradiquer** l'hétérogénéité sémantique en **fusionnant** toutes les ontologies en une ontologie globale



On peut **affronter** l'hétérogénéité sémantique en **résolvant** les problèmes au cas par cas

S'appuyant sur l'exemple de la classification chinoise, Foucault a insisté sur la diversité des points de vue taxinomiques issue de la diversité des modes de vies et des besoins : il montré que « le besoin de faire » génère dans une large mesure notre découpage catégoriel du monde. Il s'ensuit deux choses :

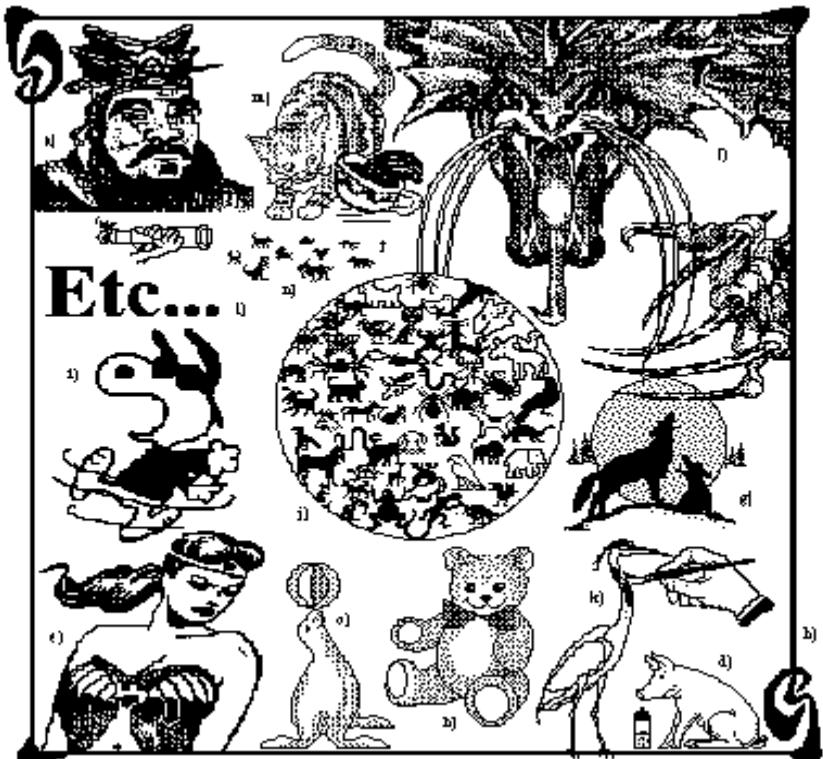
- Il se serait bien triste pour ne pas dire 'fini' un monde où régnerait une seule ontologie ;
- Faisant confiance à l'avenir il déclare le monde « ouvert » (au sens de Hewitt) et conclut à l'avènement (au sens Deleuzien) de nombreuses catégorisations « chinoises ».

Malgré cela, de nombreuses recherches ont été entreprises sur « la grande ontologie globale » comme le projet CYC de Lenat depuis le milieu des années 80 et les projets actuels sur XML – RDF – OWL ...

Dans la préface de son essai “Les mots et les Choses” (Nrf Gallimard 1966), Michel Foucault fait état d'un texte de J. L. Borges dans ses “Enquêtes” (Folio Gallimard 1974), dont la lecture dit-il “ébranle les surfaces ordonnées et tous les plans qui assagissent pour nous le foisonnement des êtres, faisant vaciller et inquiétant pour longtemps notre pratique millénaire du Même et de l'Autre” ce qui veut dire qu'il est à la fois rigolo et profond. Ce texte cite “une certaine encyclopédie chinoise” où il serait écrit que :

*Les animaux se divisent en*

*a) appartenant à l'Empereur, b) embaumés, c) apprivoisés, d) cochons de lait, e) sirènes, f) fabuleux, g) chiens en liberté, h) inclus dans cette classification, i) qui s'agitent comme des fous, j) innombrables, k) dessinés avec un pinceau très fin en poil de chameau, l) etc., m) qui viennent de casser la cruche, n) qui de loin ressemblent à des mouches.*



# Trois stratégies d'affrontement de l'hétérogénéité

7

## 1. Expression de concepts au sens des Logiques de Descriptions

- Introduction aux Logiques de Descriptions
- Tbox et requêtes hétérogènes

### Kbase Agent1

Termes : A,B,C, D, E

TBox :

$$\begin{aligned} A &\equiv E \wedge C \\ D &\equiv \forall r.C \\ B &\leq E \\ A &\leq C \end{aligned}$$

### Kbase Agent2

Termes : A,B,C, E, F,G

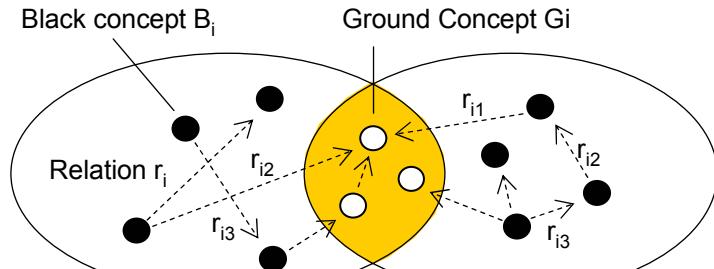
TBox :

$$\begin{aligned} F &\equiv \forall s.C \\ G &\leq C \\ E &\leq G \wedge F \\ A &\equiv B \wedge C \\ A &\leq C \end{aligned}$$

### Logiques de description

## 2. Expression sous forme de graphes de concepts lexicaux

- Le modèle de représentation WAP
- Redéfinition de concepts lexicaux
- Projection de concepts lexicaux



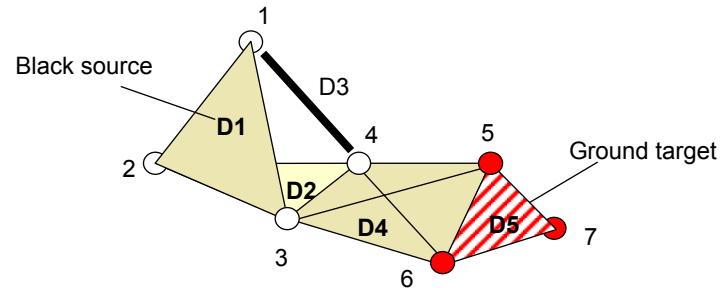
CONCEPTLIST de l'agent A

CONCEPTLIST de l'agent B

### Graphes de relations

## 3. Expression sous forme de concepts exprimés en Topologie Algébrique

- Introduction aux Complexes Simpliciaux
- Redéfinition directe de concepts
- Chemins polygonaux



### Complexes Simpliciaux

# Graphe de relations génératives

8

Approche lexicale

## Relations génératives

S = SYN[x,y] synonymie

A = ANT[x,y] antonymie

H = HYPER[x,y] hyperonymie

h = HYPO[x,y] hyponymie

K = SAMECAT[x,y] co-hyponymie

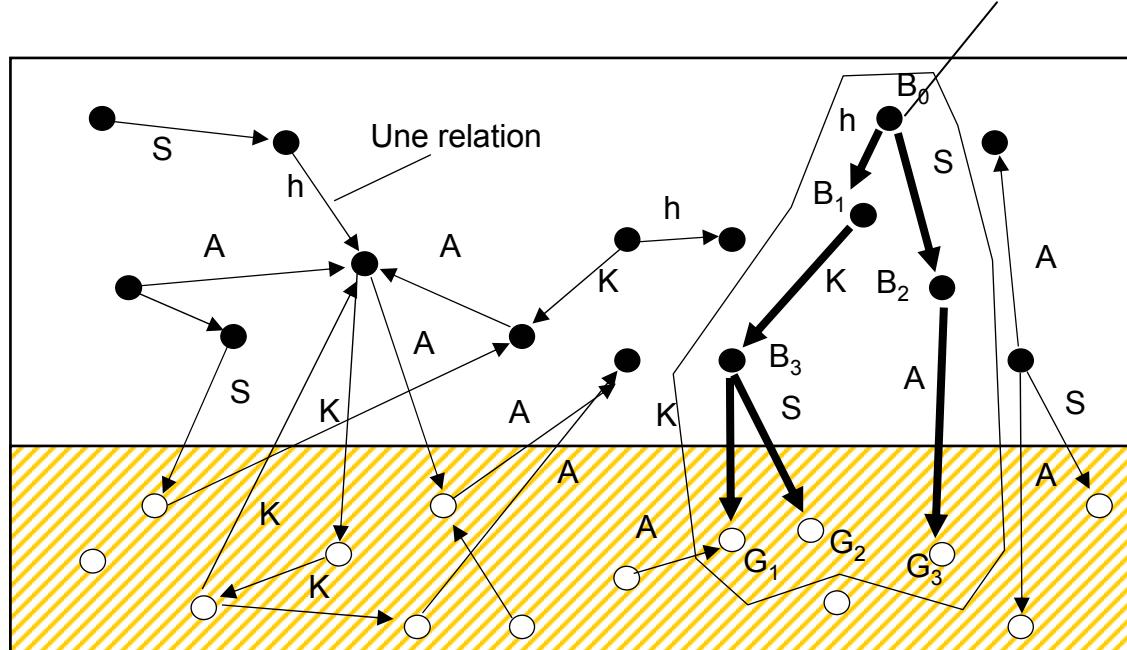
## Chemin de redéfinition

On appelle *chemin de redéfinition* (redefinition-path) d'un black concept  $B_i$ , le sous graphe du graphe des relations génératives d'une Kbase qui a pour source  $B_i$  et qui conduit (sans cycle) vers des concepts du Ground.

Redefinition-path de  $B_0$  sur  $G_1, G_2$  et  $G_3$

**Black concepts de A**

**Ground set de A**



# Exemple de projection

## Approche lexicale

On se donne une matrice aléatoire de relations pour un agent A1 (avec peu de densité) et une matrice aléatoire de relations pour un agent A2 (avec une plus forte densité). Voici les chemins de la Tbox1 qui ont au moins une image dans la Tbox2 :

```

{BEGIN[1], b[1]} → {{BEGIN[1], b[7]}, {BEGIN[1], b[10]}}
```

```

{BEGIN[2], a[4], b[2]} → {{BEGIN[2], b[9], a[14]}, {BEGIN[2], b[16], a[4]}, {BEGIN[2], b[16], a[6]}, {BEGIN[2], b[16], a[7]}}
```

```

{BEGIN[2], a[4]} → {{BEGIN[4], a[3]}, {BEGIN[4], a[4]}, {BEGIN[4], a[11]}}
```

```

{BEGIN[2], c[6]} → {{BEGIN[6], c[1]}, {BEGIN[6], c[5]}, {BEGIN[6], c[8]}}
```

```

{BEGIN[3], a[5]} → {{BEGIN[5], a[1]}, {BEGIN[5], a[3]}}
```

```

{BEGIN[4], b[2]} → {{BEGIN[2], b[9]}, {BEGIN[2], b[16]}}
```

```

{BEGIN[4], b[2], a[4]} → {{BEGIN[4], a[3], b[11]}, {BEGIN[4], a[11], b[8]}}
```

```

{BEGIN[4], b[2], c[6]} → {{BEGIN[6], c[1], b[7]}, {BEGIN[6], c[1], b[10]}, {BEGIN[6], c[8], b[5]}}
```

```

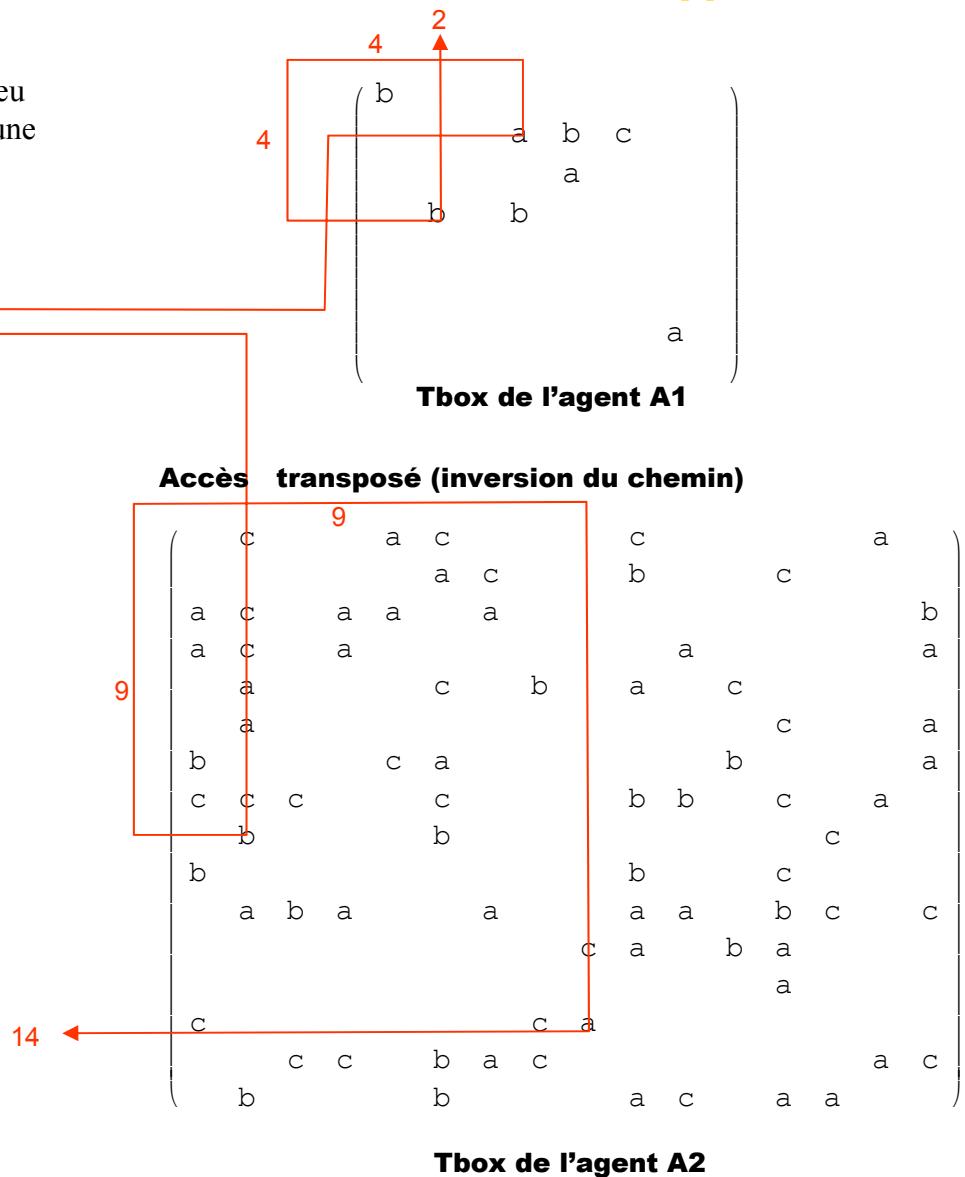
{BEGIN[7], a[7]} → {{BEGIN[7], a[3]}, {BEGIN[7], a[11]}, {BEGIN[7], a[15]}}
```

Total number of paths in Tbox1: 12  
Number of paths of Tbox1 with at least an image in Tbox2: 9  
Number of paths of Tbox1 with no image in Tbox2: 3

Le début d'un chemin

- dans Tbox1 est noté BEGIN[Blackconcept],
- dans Tbox2 est noté BEGIN[Groundconcept].

La notation  $r[i]$  indique que par la relation  $r$  on va vers le concept  $i$ .



# Probabilité d'existence d'un chemin de redéfinition

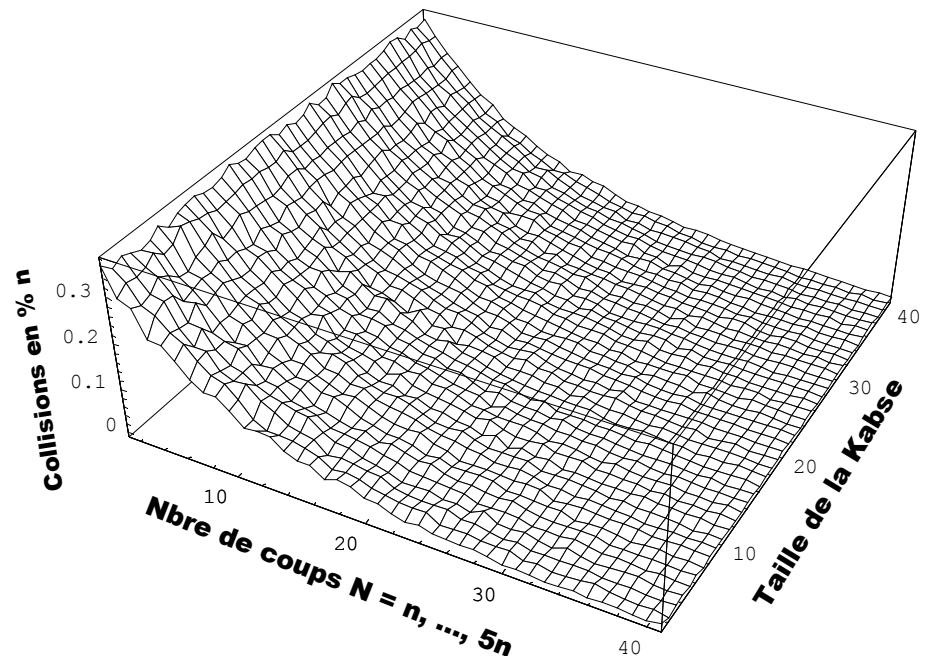
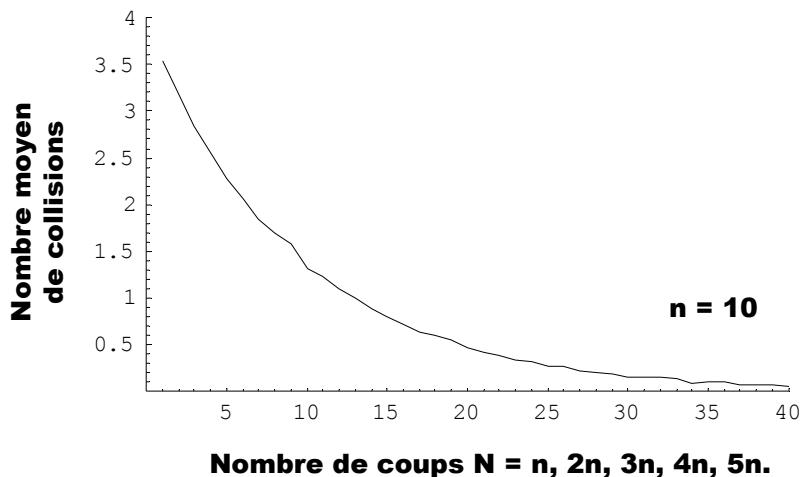
10

## Approche lexicale

Le problème de l'existence d'au moins un chemin relationnel entre deux concepts i,j quelconques d'une Kbase peut être formalisé par la probabilité de tirer au sort une matrice diagonale de relation non vide ou encore toute permutation de cette matrice :

$$\begin{pmatrix} \mathbf{r} & \square & \square & \square & \square \\ \square & \mathbf{r} & \square & \square & \square \\ \square & \square & \mathbf{r} & \square & \square \\ \square & \square & \square & \mathbf{r} & \square \\ \square & \square & \square & \square & \mathbf{r} \end{pmatrix} \quad \dots \quad \begin{pmatrix} \mathbf{r} & \square & \square & \square & \square \\ \square & \square & \mathbf{r} & \square & \square \\ \square & \square & \square & \square & \mathbf{r} \\ \square & \mathbf{r} & \square & \square & \square \\ \square & \square & \square & \mathbf{r} & \square \end{pmatrix}$$

La courbe ci-dessous donne le nombre moyen de collisions de relations sur une colonne en fonction du nombre de relations tirées  $N$  : cette probabilité commence vers 3.5 pour  $N = n$  et tend vers 0.



Le graphique ci-dessus ajoute la variation du paramètre de taille de la Kbase pour  $n = 10, \dots, 50$  concepts. Le nombre moyen de collisions est normalisé par rapport avec la taille  $n$  de la Kbase (il est exprimé en % de la taille et non en valeur absolue comme à gauche). On constate 1) le modèle statistique se comporte de manière homogène 2) dès que  $N > 3.5n$ , la probabilité d'obtenir une matrice 'percolante' (i.e. qui inclut au moins une permutation de la matrice diagonale) est de 1, voulant dire que tout couple, i,j est relié par au moins un chemin de redéfinition.

### Cellule

Les cellules sont les briques de base pour la construction d'un espace CAT:

« *On appelle K-cellule, un morceau d'espace abstrait de dimension k et qui ne contient aucun trou.* »

### Complexe

Un complexe cellulaire est un ensemble de cellules qui vérifie :

- Le bord d'une cellule est constitué d'une union finie de cellules ;
- L'intersection de deux cellules es soit vide soit une autre cellule.

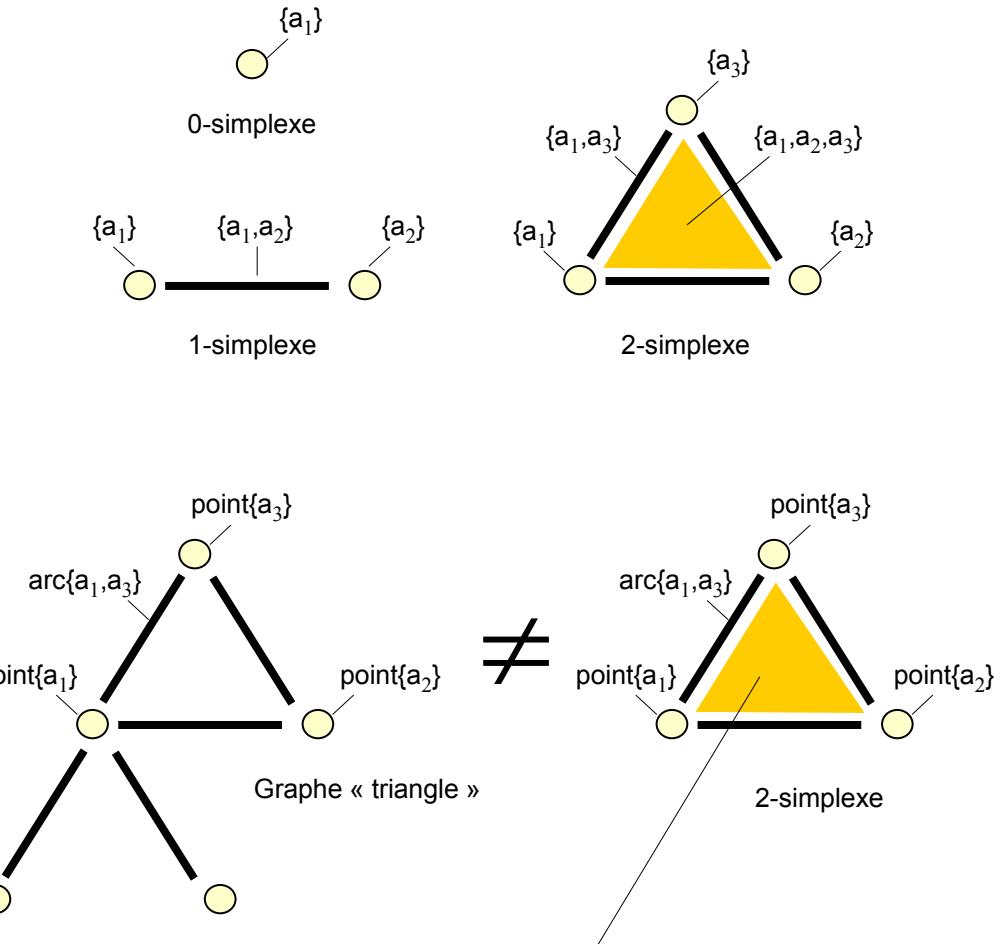
« *Un complexe simplicial abstrait est un couple  $(V, C)$  où  $V$  est un ensemble de sommets et  $C$  un ensemble de parties finies de  $V$  telles que si  $s \in C$  alors toutes les parties  $s' \in s$  appartiennent aussi à  $C$*  »

### Simplexe

Les éléments de  $C$  sont des *simplexes* abstraits.

La dimension d'un simplexe est son nombre de sommets moins 1.

La dimension d'un complexe est celle du plus grand simplexe qu'il contient.



la surface  $\{a_1, a_2, a_3\}$  peut exprimer une relation entre trois sous-structures :  $\text{arc}\{a_1, a_2\}$ ,  $\text{arc}\{a_1, a_3\}$ ,  $\text{arc}\{a_2, a_3\}$ .

## **Redéfinition directe de frames à un niveau**

## Redéfinition directe de concepts

On se donne la Tbox suivante où les domaines ne sont pas explicités car ce sont les structures composées simples, i.e. à un seul niveau de field (ou frames) qui nous intéressent ici :

```

(* some domains defs *)
D1=def1,D2=def2,D3=def3,
D4=def4,D5=def1
(* one-level frames *)
Tx = {T1,T2}
Ty = {T1,T2,T3}
Tz = {T1,T2,T3,T4,T5}
(* fields *)
T1=D1,T2=D2,T3=D3,
T4=D4,T5=D5

```

## **Algorithme** (G est le Ground)

### Phase 1 : niveau des fields $T_i = D_i$

R1  $T_i \in G$  et  $D_i \in G$  : pas de pb

R2  $T_i \notin G$  et  $D_j \in G$

$\exists Tg \in G \mid Tg=Di$  alors  $Ti$  isako( $Tg$ )

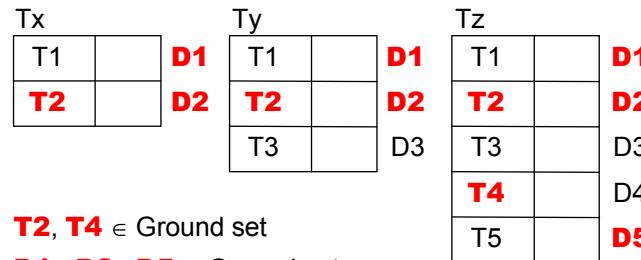
$\neg \exists Tg \in G \mid Tg = Dj \text{ alors } Ti \text{ isako}(Dj)$

R3  $T_i \in G$  et  $D_i \notin G$  erreur (supposé traité)

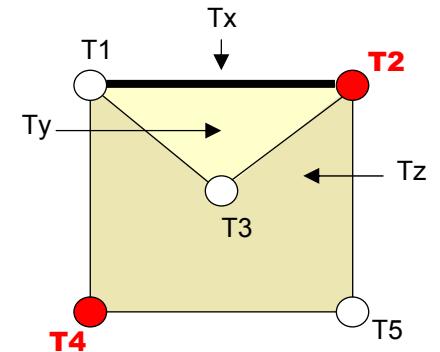
R4  $T_i \notin G$  et  $D_i \notin G$  redéfinir  $D_i$  puis goto 2 b

Phase 2 : au niveau des frames  $T_j \in \{T1, T2, \dots\}$

Les frames jouent alors le rôle des valeurs pour les domaines dans l'algorithme déjà vu qui est utilisé.



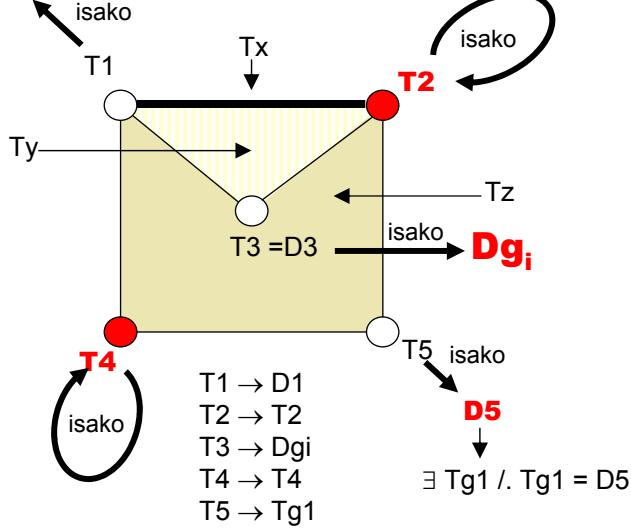
## Structure du WAP



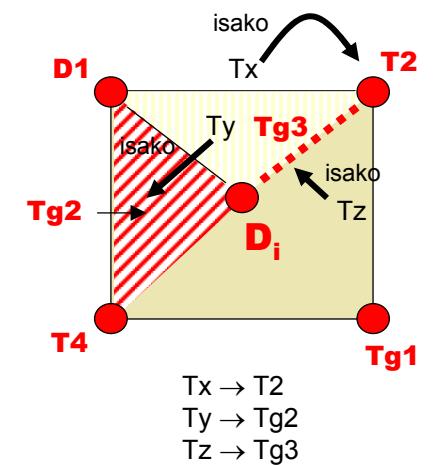
## Representation simpliciale

$\neg \exists \text{ Tgi} / . \text{ Tgi} = \text{D1}$

**D1**



## 1. Field level redefinition

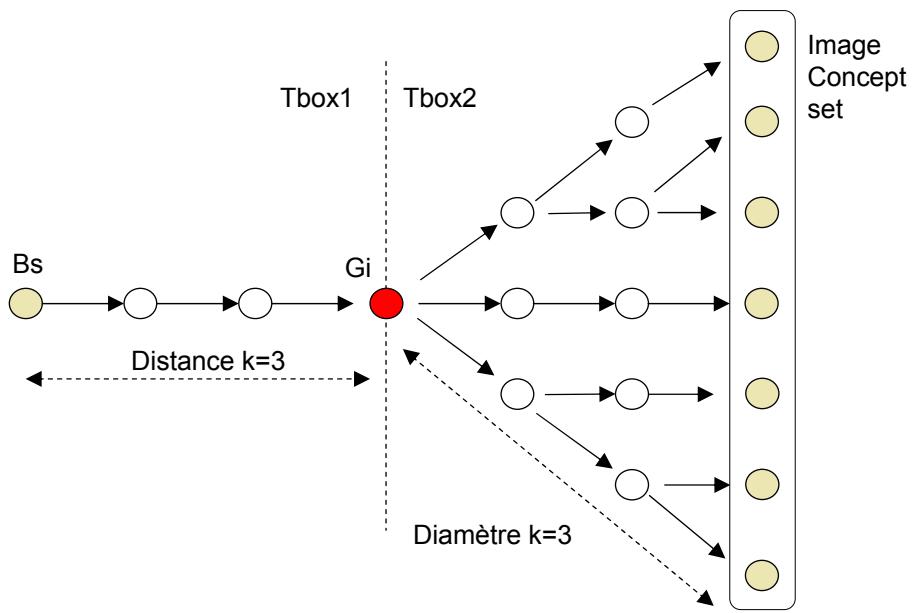


## 2. Frame level redefinition

# Comparaison entre les graphes et les complexes

13

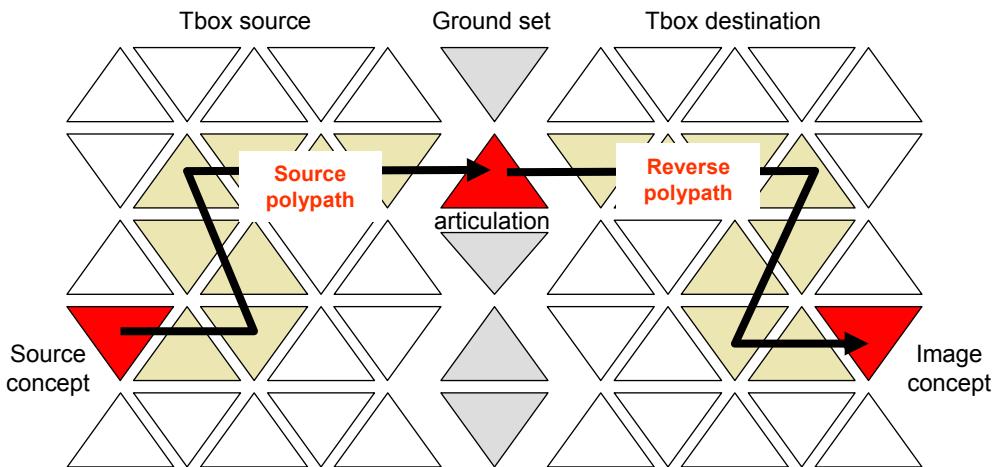
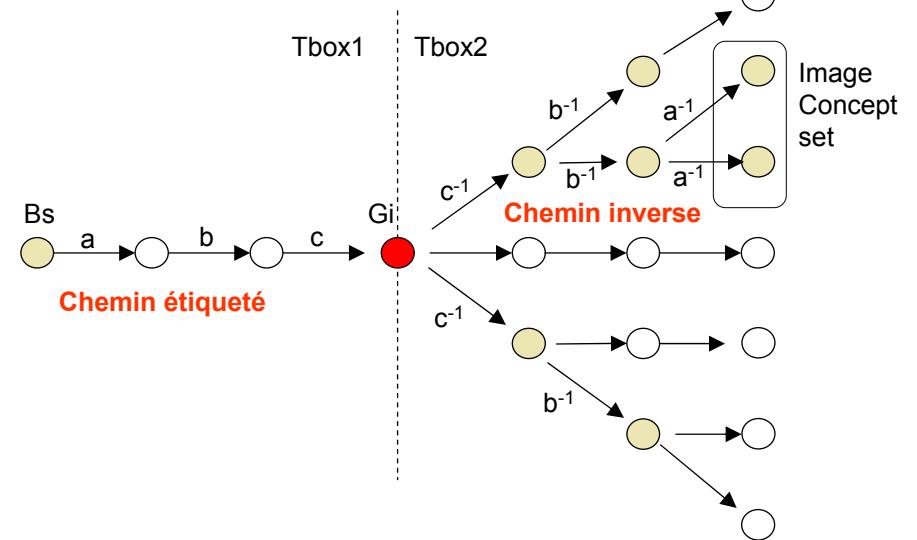
## Chemins polygonaux



Dans l'approche à base de graphes, le chemin de redéfinition est monodimensionnel.

- Si on ne tient pas compte des relations (figure ci-dessus) le chemin est analogue à un chemin d'accès de fichier : le calcul de l'image inverse produit un grand nombre d'images possibles,
- Si on prend en considération les nom des relations (figure en haut à droite) le chemin est plus discriminant et l'image inverse est beaucoup plus précise.

Dans l'approche à base de complexes simpliciaux, le chemin de redéfinition est multidimensionnel (pour faciliter la lecture, dans la figure de droite, tous les simplexes du complexe où on navigue sont de dimension 3 – ce sont des faces). Dans ce cas, l'image du chemin polygonal est beaucoup plus précise et surtout elle tient compte, à chaque pas, des relations non en termes de simples identifiants atomiques (a,b,c,...) mais en termes de leur vraie extension.



## 1. Modéliser l'Hétérogénéité Sémantique

- Qu'est-ce que l'Hétérogénéité Sémantique ?
- Modélisation en Logiques de Descriptions : Tbox et Abox hétérogènes

## 2. Expression sous forme de graphes de concepts lexicaux

- Le modèle de représentation WAP
- Redéfinition de concepts lexicaux
- Projection de concepts lexicaux

## 3. Expression sous forme de concepts exprimés en Topologie Algébrique

- Introduction aux Complexes Simpliciaux
- Redéfinition directe de concepts
- Chemins polygonaux



English

[Accueil](#)[Animation](#)[Recherche](#)[Publications](#)[Démos](#)[InterViews](#)[Référence](#)[Grasp](#)[MLK](#)[CommonSense](#)[WeakAliens](#)[Daft](#)[Zilla](#)[Buzz](#)[Magenta](#)[ASP navigation](#)[Enseignement](#)[Stages 2004](#)[Cours en ligne](#)[Paniers Agents](#)

Groupe AMI

IMCI

## WEAK ALIENS Semantic heterogeneity between informational agents

In the framework of the Internet, a new crucial problematics arises: how can we deal with the semantic heterogeneity issuing from the abundant diversity of informational components that are produced at different times and locations by different people with different aims in mind? XML DTDs approaches lead to the multiplication of committees for the standardization of ontologies but we know since Borges and Foucault that the "Great Ontology" is an utopia, at least as far as the real world is concerned [See the [Chinese Ontology](#) of J.L. Borges].

Instead, in this study, we take the opposite position: assuming that semantic heterogeneity is the basic rule, we claim that informational agents are not complete "aliens" to each other (thus the term weak aliens) and that they often share some semantic concepts on which we can rely to transfer the semantics of unknown entities from a knowledge base to another.

The term "aliens" was inspired by the paper of Marvin Minsky, published in "Extraterrestrials: Science and Alien Intelligence", (Edward Regis, Ed.) Cambridge University Press 1985 [ [available in html](#) ].

— The home page of [Erika Valencia](#)

### Publications

- J.-P. Sansonnet, E. Valencia, [Q-chaines : Formalisme Simplicial pour L'interaction entre Agents Sémantiquement hétérogènes](#), poster session, JNMR'03, Paris, 27-29 nov 2003 [ [poster](#) ]
- E. Valencia, J.-P. Sansonnet, [Modélisation de l'Hétérogénéité Sémantique entre Agents Dialogiques](#), JFIADSMA 2000, St Etienne, oct 2000.



The Shadoks by J. [Roussel](#) are some kind of funny Aliens with an odd physics and philosophy, speaking a daft language: the Gabuzomeu. But, we humans, share so much with them....