



**Séminaire de l'équipe CRAC – 9 mars 2017**

# Une ontologie pour l'enseignement des statistiques : défis et perspectives

**Jean-Marc Meunier, MCF, Laboratoire  
Paragraphe – CRAC, Université Paris 8**

# Les préconceptions en statistiques

# La loi des grands nombres et échantillonnage

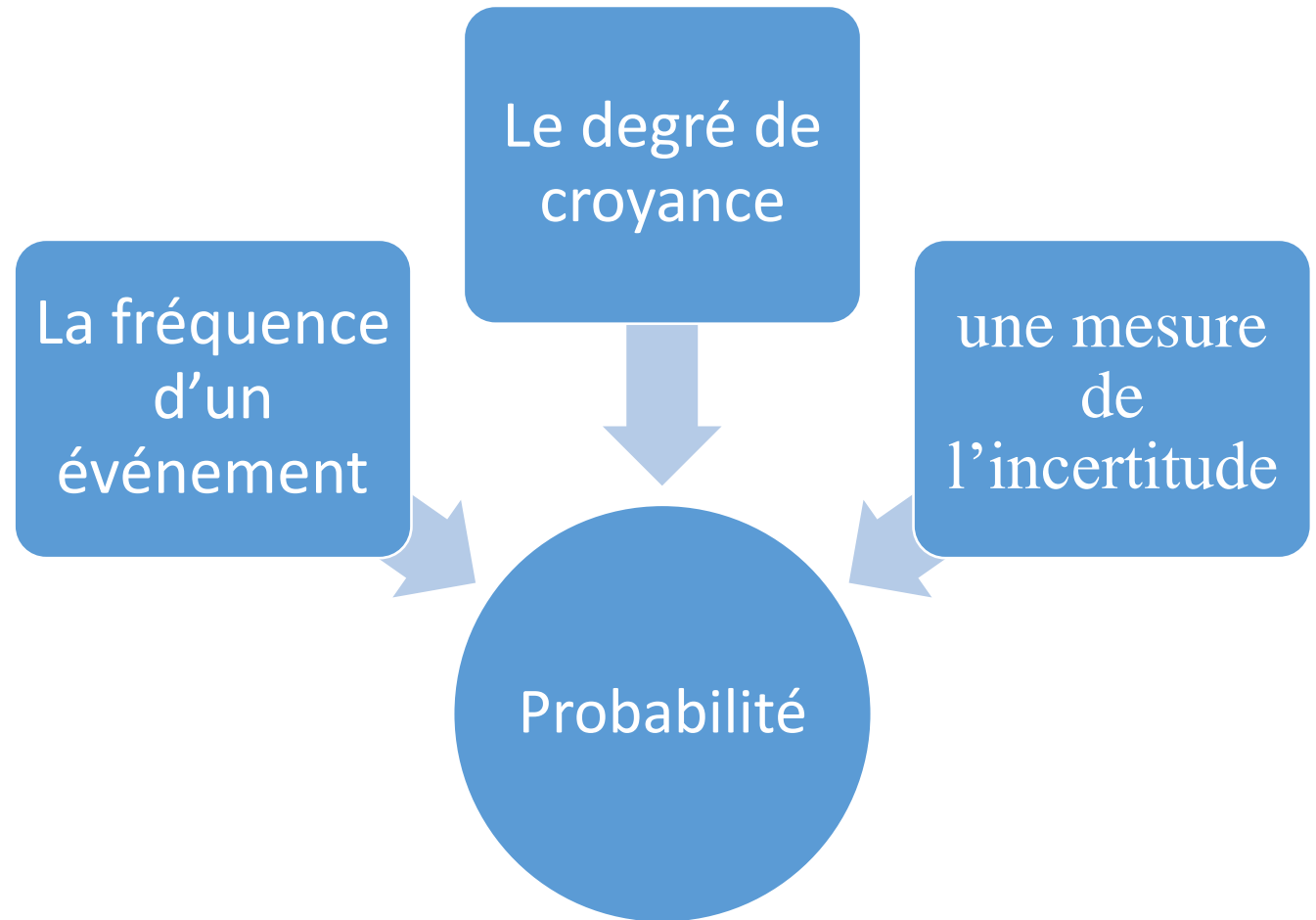
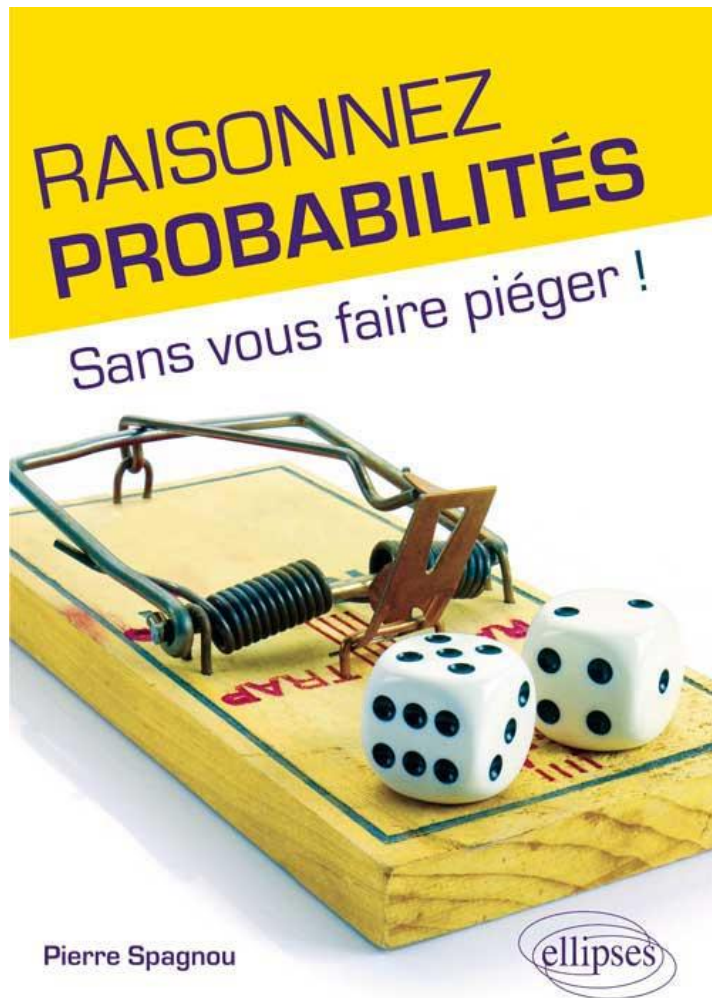
Loi des grands nombres : plus la taille d'un échantillon est grande, plus ses caractéristiques tendent à se rapprocher de celles de la population



Le problème de la maternité (Kahneman et Tversky, 1972)

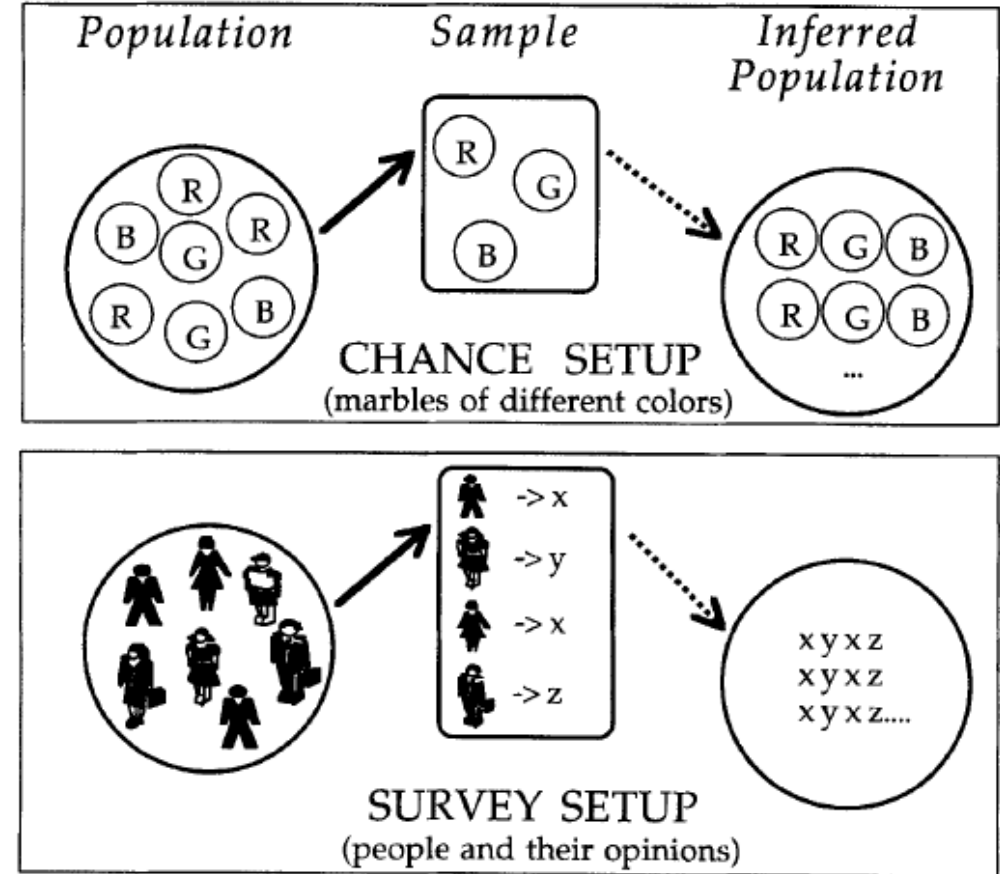
- Les jours où on observe que 60% des naissances sont des garçons sont-ils plus fréquents dans un hôpital où il y a 15 naissances que dans un hôpital où il y a 45 naissances par jour
- Le QI de la population est connu pour être en moyenne de 100. Vous avez sélectionné un échantillon de 50 sujets. Le premier enfant testé à un QI de 150. Que peut-on prédire au sujet de la moyenne de l'échantillon ?

# Trois conceptions de la probabilité :(Well, Pollatsek & Konold, 1983; Konold, 1989)



# Concepts quotidiens vs scientifique

- L'échantillonnage : approche singulière ou distributionnelle (Schwartz et Goldman, 1996)
- Comprendre l'espace des échantillons et la variabilité (Nisbett (1983))
- Langage quotidien et langage scientifique (Mayen & Diaz, 2010; Nacarato, 2014; Taabane, 2006; Dunn et Al, 2016; Lesser et al; 2009 )



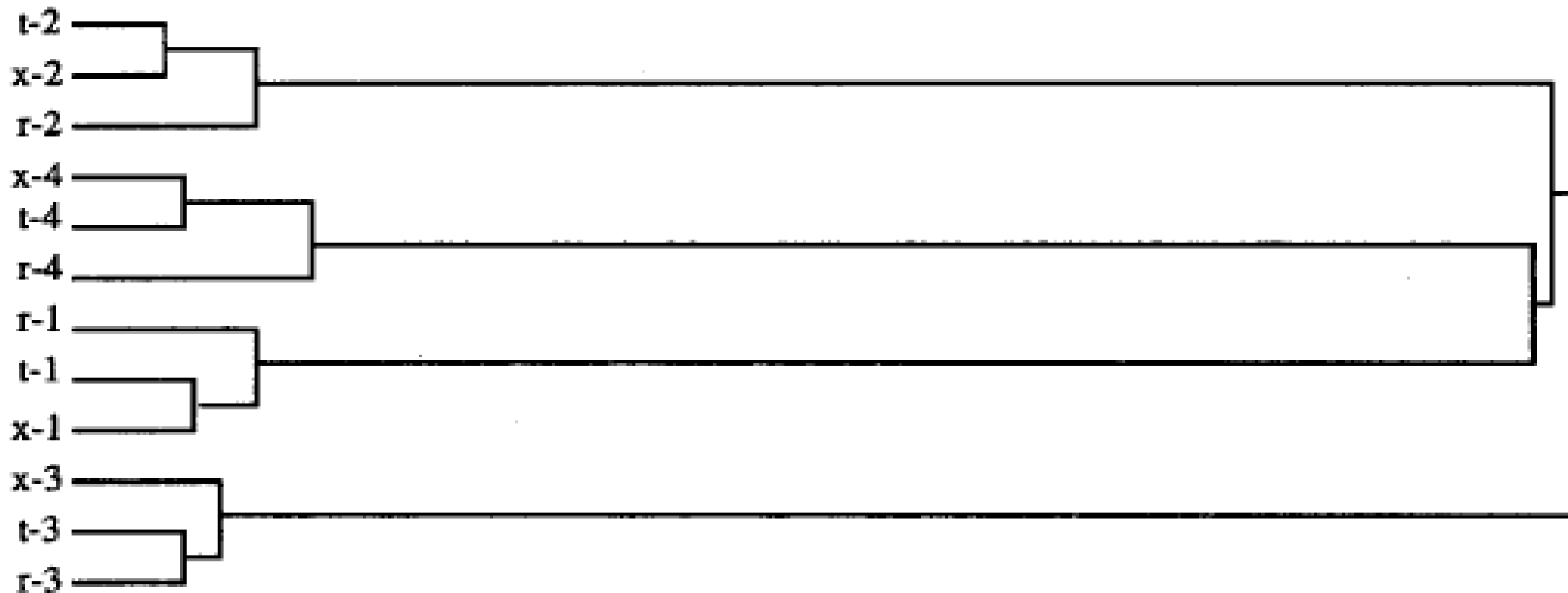
# Identifier les procédures : un challenge

- L'effet des caractéristiques de surface (Quilici et Mayer, 1996)
  - Catégorisation de 12 problèmes
  - Quatre domaines croisés avec trois types de test
- Résultats
  - Sans exemple préalable : catégorisation sur le contenu
  - Avec exemples préalables; catégorisation sur le type de test

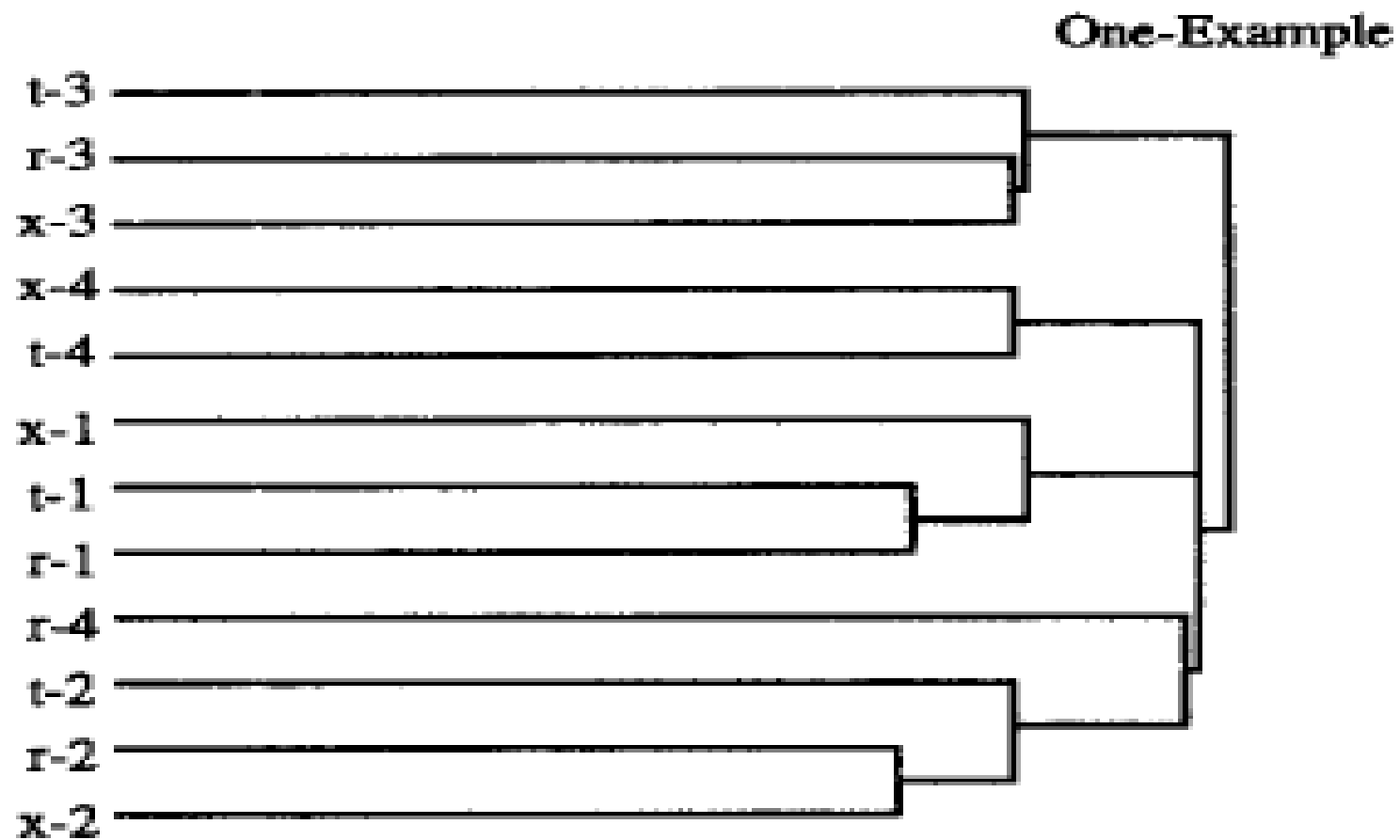


# Identifier les procédures : un challenge

No-Example



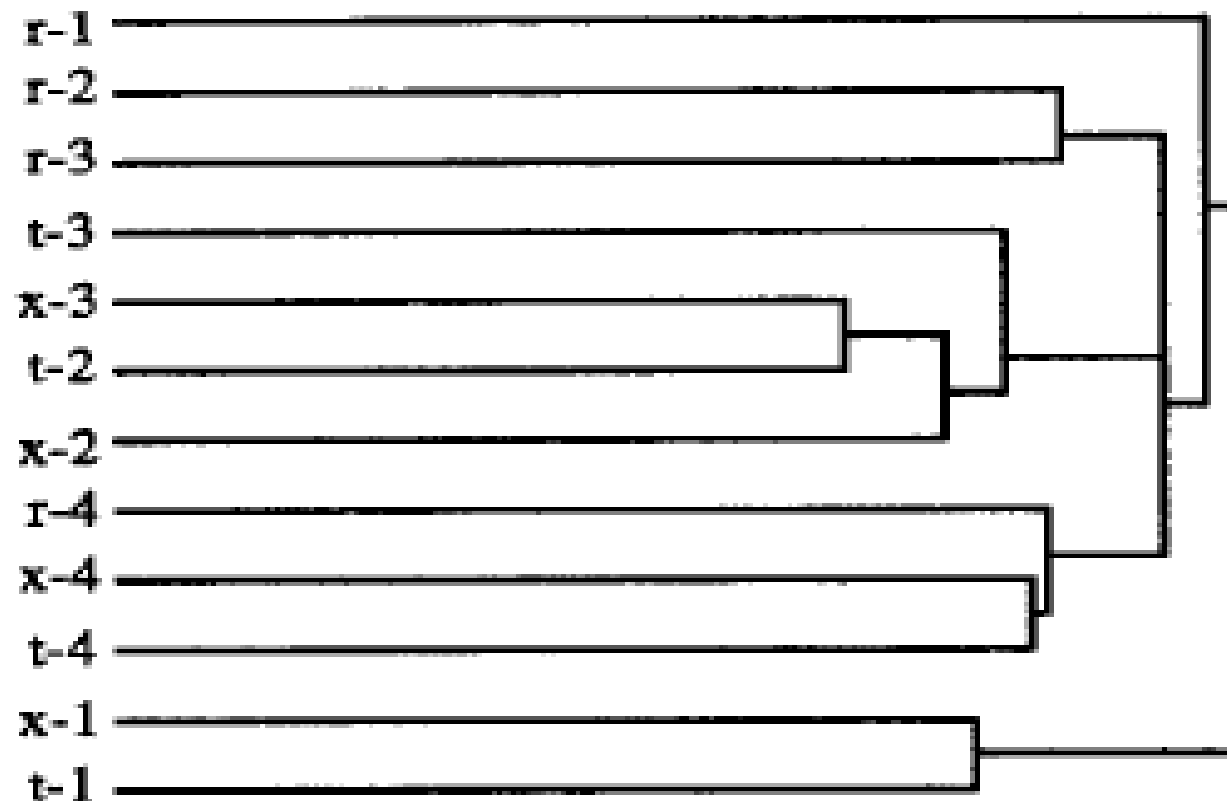
# Identifier les procédures : un challenge





# Identifier les procédures : un challenge

Three-Example



# Identifier les procédures : un challenge

- Identification difficile de la procédure adéquate (Gardner et Hudson, 1999)
  - 34 énoncés
  - Quelle procédure ?  
Confiance ? familiarité ?
  - Performance médiocre
  - Confiance faiblement corrélée à la performance
  - Confiance fortement corrélée à la familiarité

P Correct	Undergraduate	Fourth-year	Master	PhD
0,5		4		
0,4	2		1	
0,3	1	2		
0,2	2			
0,1	3	1		1
0,0	4		2	
Total	12	7	3	1

# Identifier les procédures : un challenge

- Le rôle de l'explicitation de la structure vs schéma de résolution (Quilici et Mayer, 2002)



Group	Sorting analysis	
	Mean sorting score	Proportion of students who sorted by structure
<i>Preliminary study</i>		
Pretest	-0.18	25/51
Posttest	+0.11	36/51
<i>Main study</i>		
No aids	-0.09	16/36
Schema aids	+0.23	28/36
Example aids	+0.38	30/36
Both aids	+0.24	28/36

# Une structuration non spontanée (Roberts, 1999)

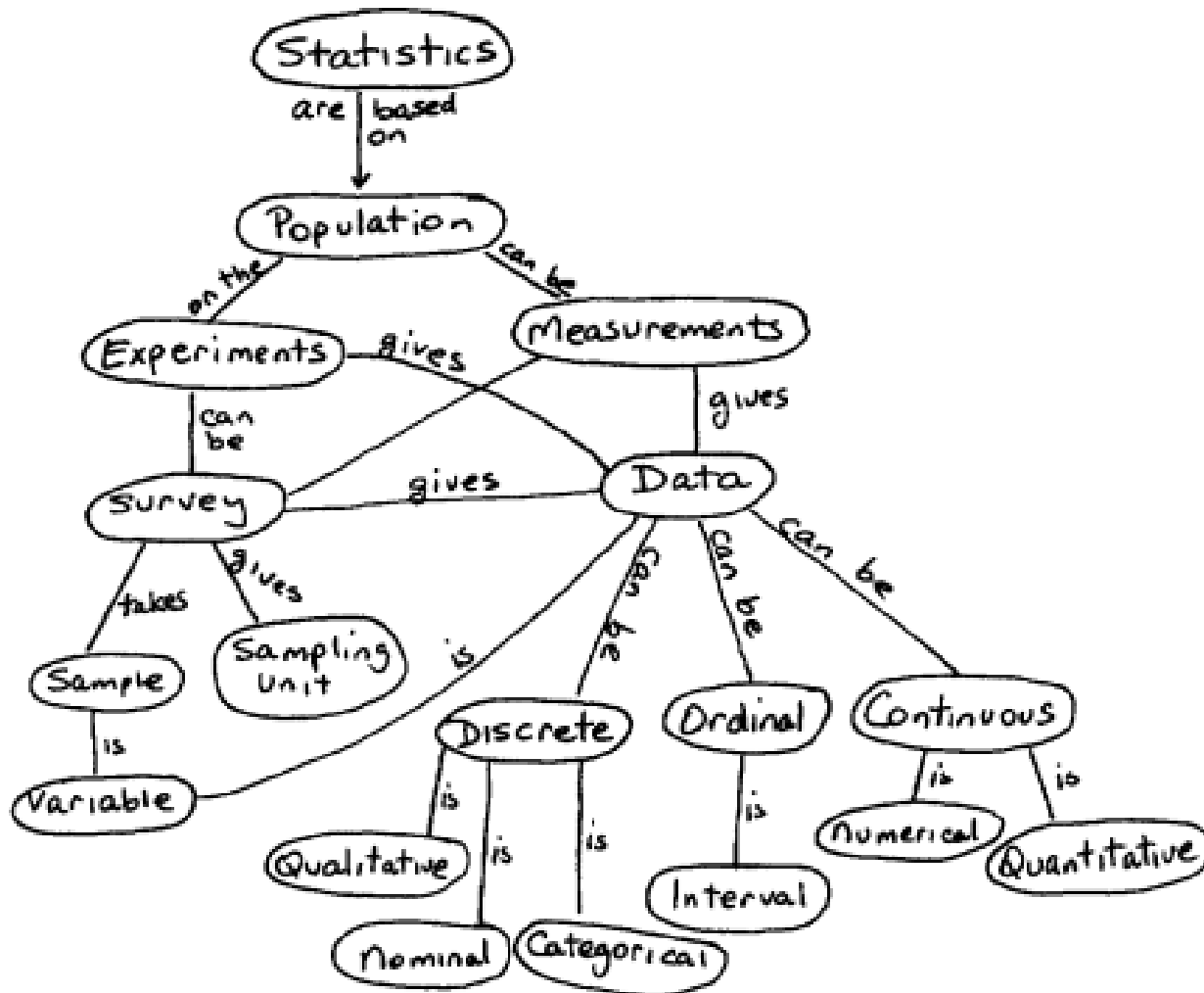


Figure 2. Betty's map.

- Nécessite un apprentissage
- Relations hétérogènes
- Comporte de nombreuses erreurs
- difficiles à remettre en cause

# Application à l'enseignement (Chiou, 2009)

- 51 étudiants 1er cycle économie et gestion répartis dans trois groupes
  - Control
  - Collaborative map
  - Individual map
- Méthodologie pré et post test (10 QCM)

# Application à l'enseignement (Chiou, 2009)

Table 5. Analysis of variance in post-test scores.

Source of variance	SS	Df	MS	<i>F</i>
Treatments	6893.81	2	3446.90	14.702**
Error	11253.88	48	234.46	
Bonferroni post hoc comparison				
Groups	Difference in means	Standard error	<i>p</i>	
ICMING-TTE	14.82	5.25	0.021*	
CCMING-TTE	28.47	5.25	0.000**	
CCMING-ICMING	13.65	5.25	0.037*	

Note: \* $p < 0.05$ ; \*\* $p < 0.01$ .

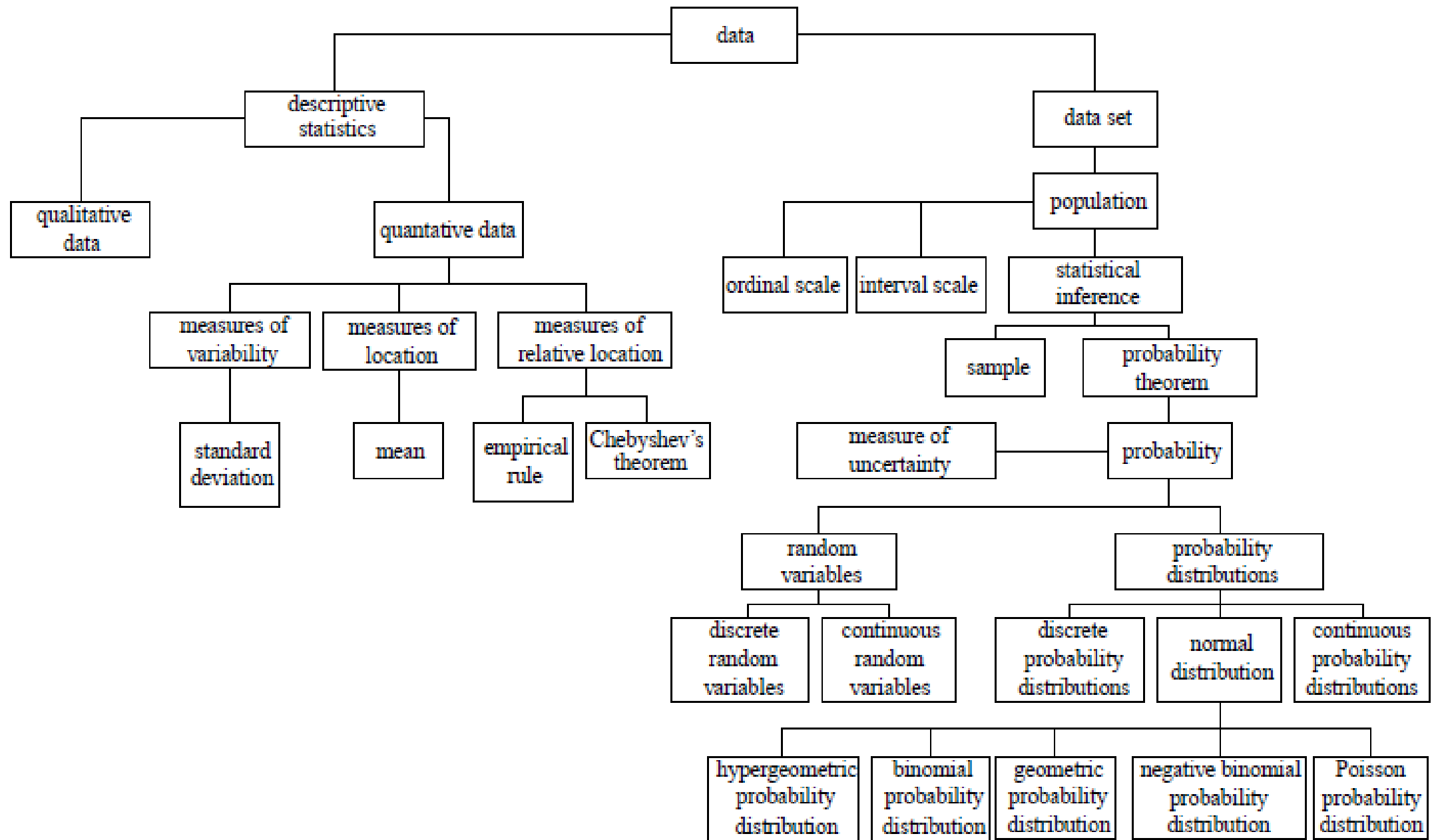


Figure 1. Student CM before correction by the teacher.

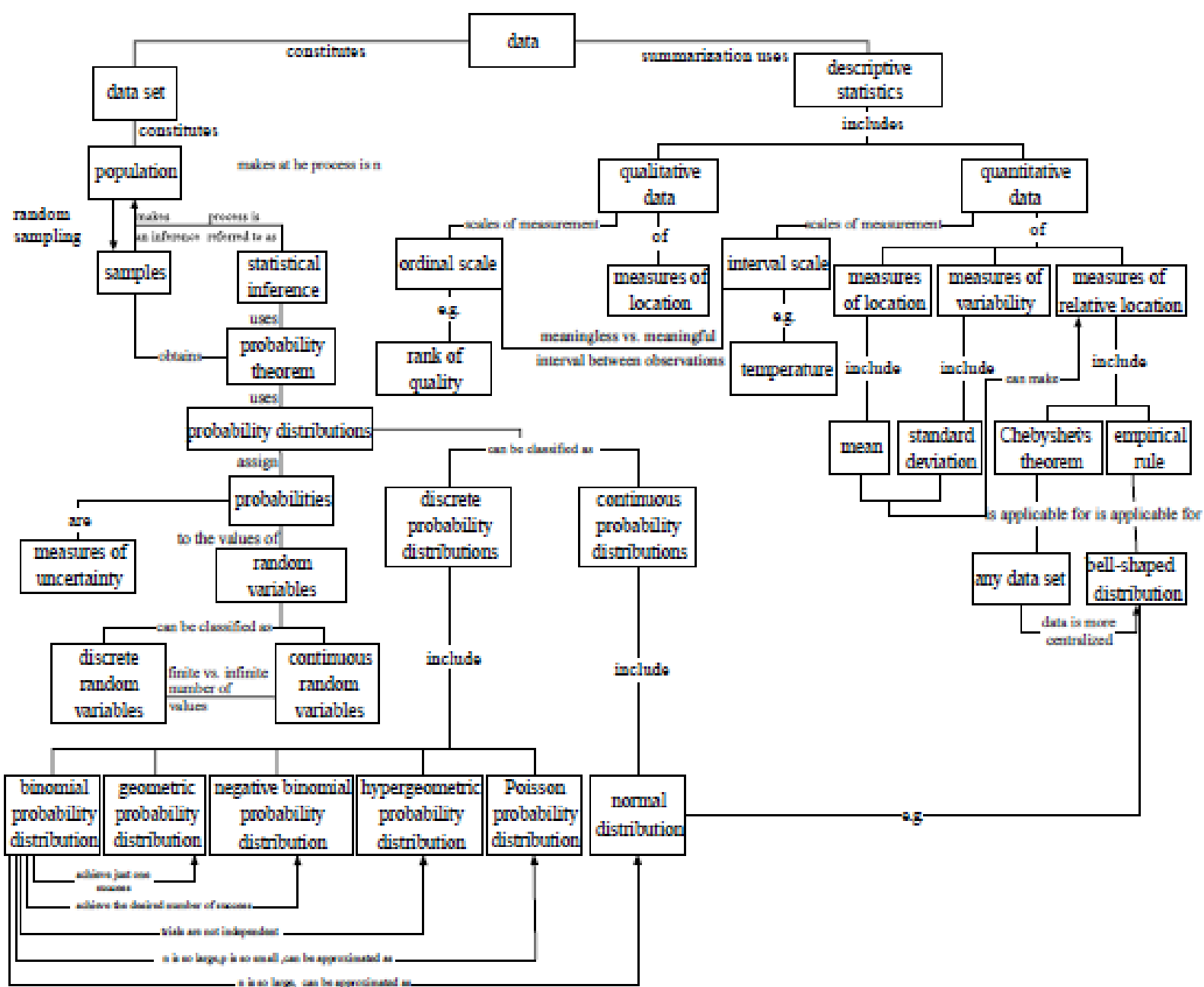


Figure 2. Student CM after correction by the teacher.



# Evaluation à l'aide des CM (Yin, 2012)

- Distinction entre tree diagrams et concept maps
- Usage pour l'évaluation.
  - Construire un arbre à partir de rien (tree diagram task)
  - Ranger les concepts (arranging the vertices)
  - Remplir les noeuds

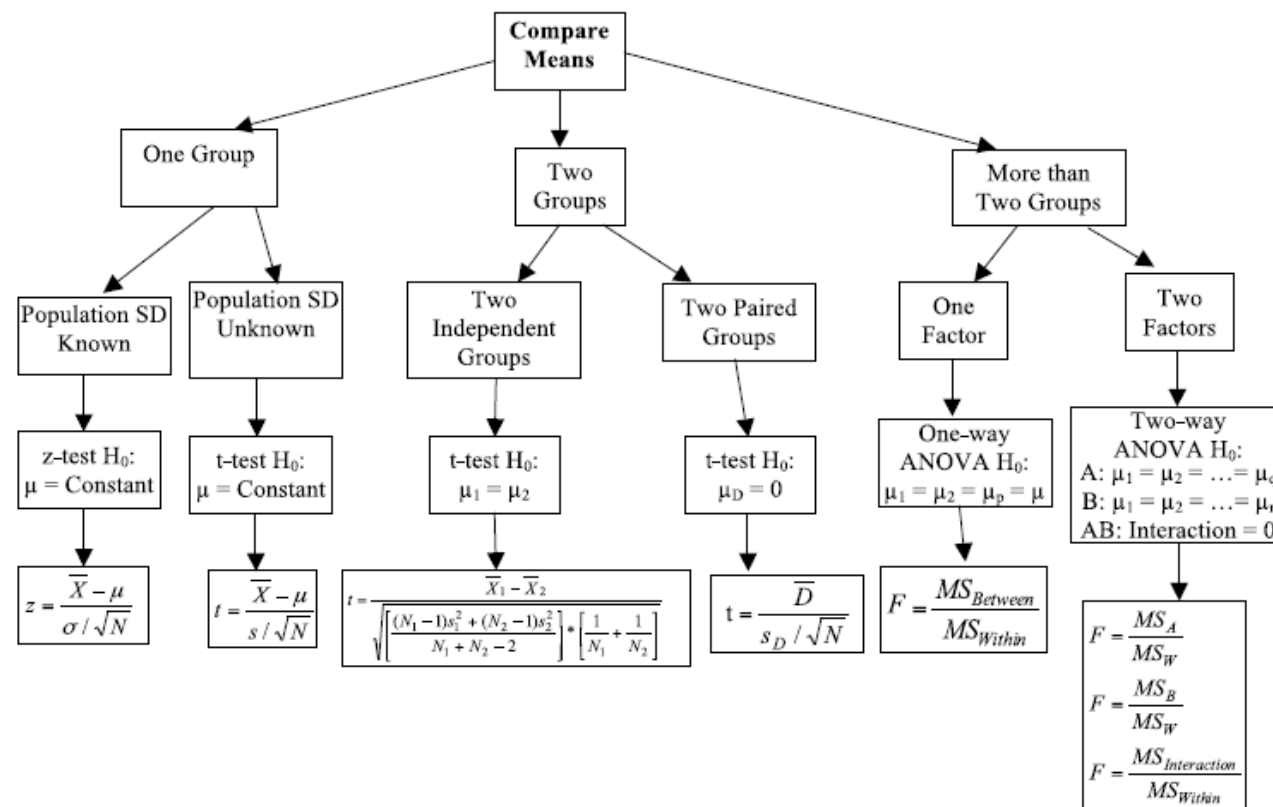


FIGURE 1 Tree diagram for comparing means.

# Evaluation à l'aide des CM (Yin, 2012)

TABLE 1  
Descriptive Analysis of the Students' Responses to the Survey

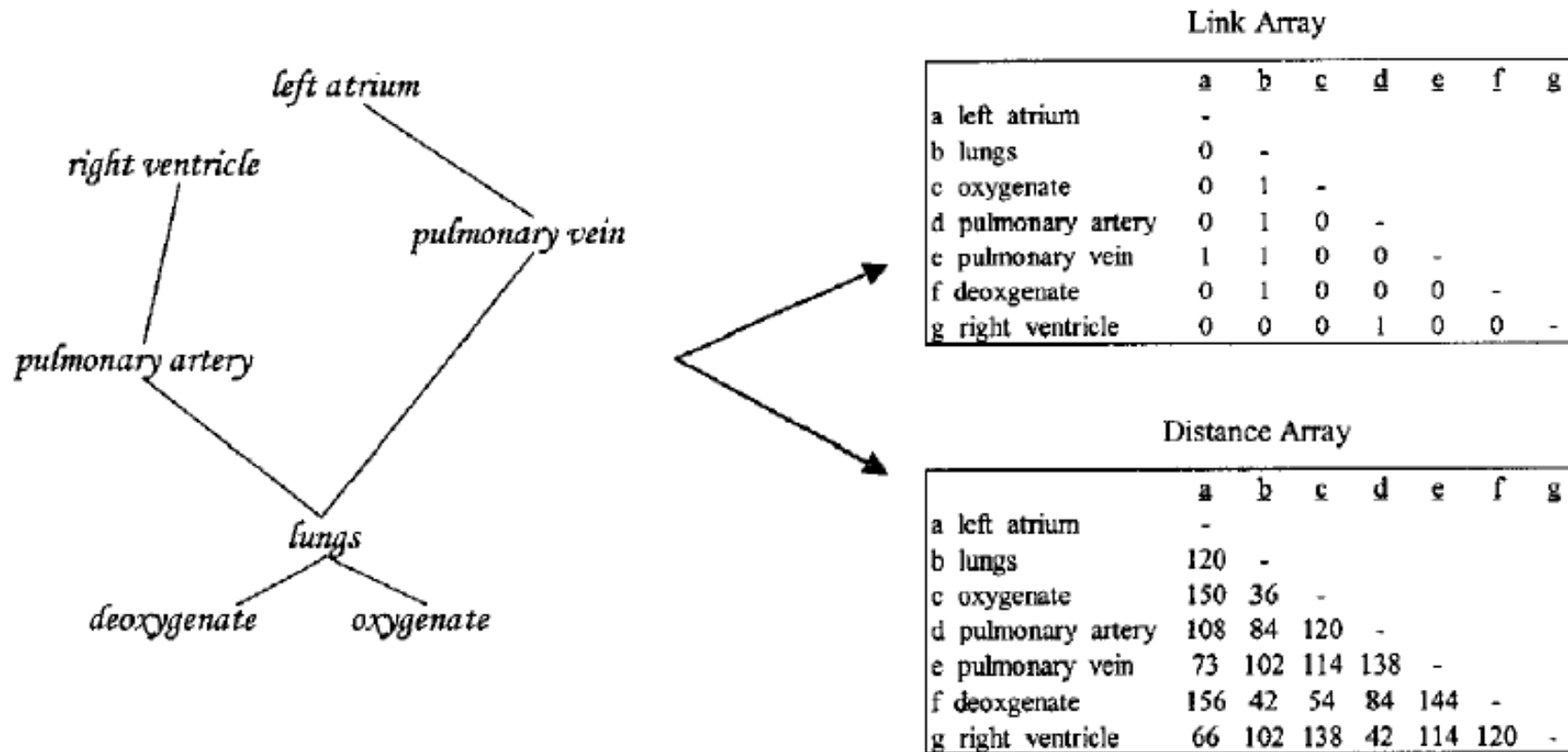
<i>The Graphic Organizer<sup>a</sup> Helps Me</i>	<i>Relative Frequency</i>					<i>Descriptive</i>	
	<i>Strongly Disagree</i>	<i>Disagree</i>	<i>Neutral</i>	<i>Agree</i>	<i>Strongly Agree</i>	<i>M</i>	<i>SD</i>
1. Understand these concepts.	0.0	5.0	22.5	62.5	10.0	3.78	.70
2. Memorize these concepts.	0.0	15.0	30.0	42.5	12.5	3.53	.91
3. Solve problems that are related to these concepts.	0.0	10.0	22.5	47.5	20.0	3.78	.89
4. Reflect on what I have learned.	0.0	7.5	25.0	45.0	22.5	3.83	.87
5. Identify what I have not learned well.	0.0	10.0	32.5	35.0	22.5	3.70	.94
6. See relationships among different concepts.	0.0	2.5	10.0	50.0	37.5	4.23	.73
7. Organize pieces of separate information into a coherent way.	0.0	0.0	25.0	47.5	27.5	4.03	.73
8. See the big picture of what we have learned.	0.0	5.0	12.5	50.0	32.5	4.10	.81
9. Graphic organizers should be used in introductory statistics classes	2.5	2.5	25.0	50.0	20.0	3.83	.87

*Note.*  $N = 36$ .

<sup>a</sup>A graphic organizer was used in the survey rather than a tree diagram. However, for consistency in this article, the tree diagram was used in the corresponding discussion.

# Evaluation automatique des CM (Taricani et Claiana, 2006)

Figure 1 □ The link and distance data arrays of a simple map.

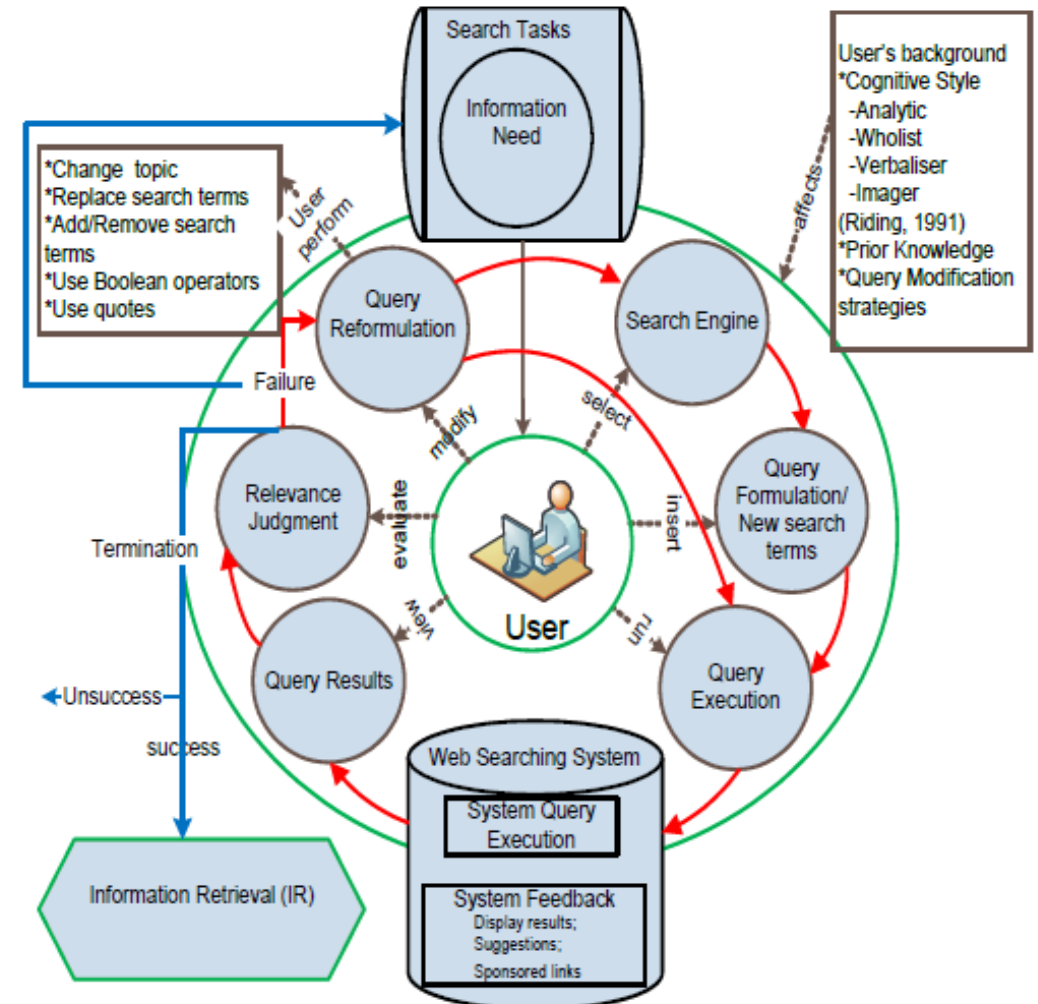


# Apprendre à l'heure d'internet



# Des compétences à chercher l'information (Kinley, Tjondronegoro, & Partridge, 2010)

- Stratégie globale vs analytique
- Importance de l'expérience
- Importance des connaissances
- Décalage possible avec les systèmes d'indexation



# Le projet Ontostats

# Partenaires du projet



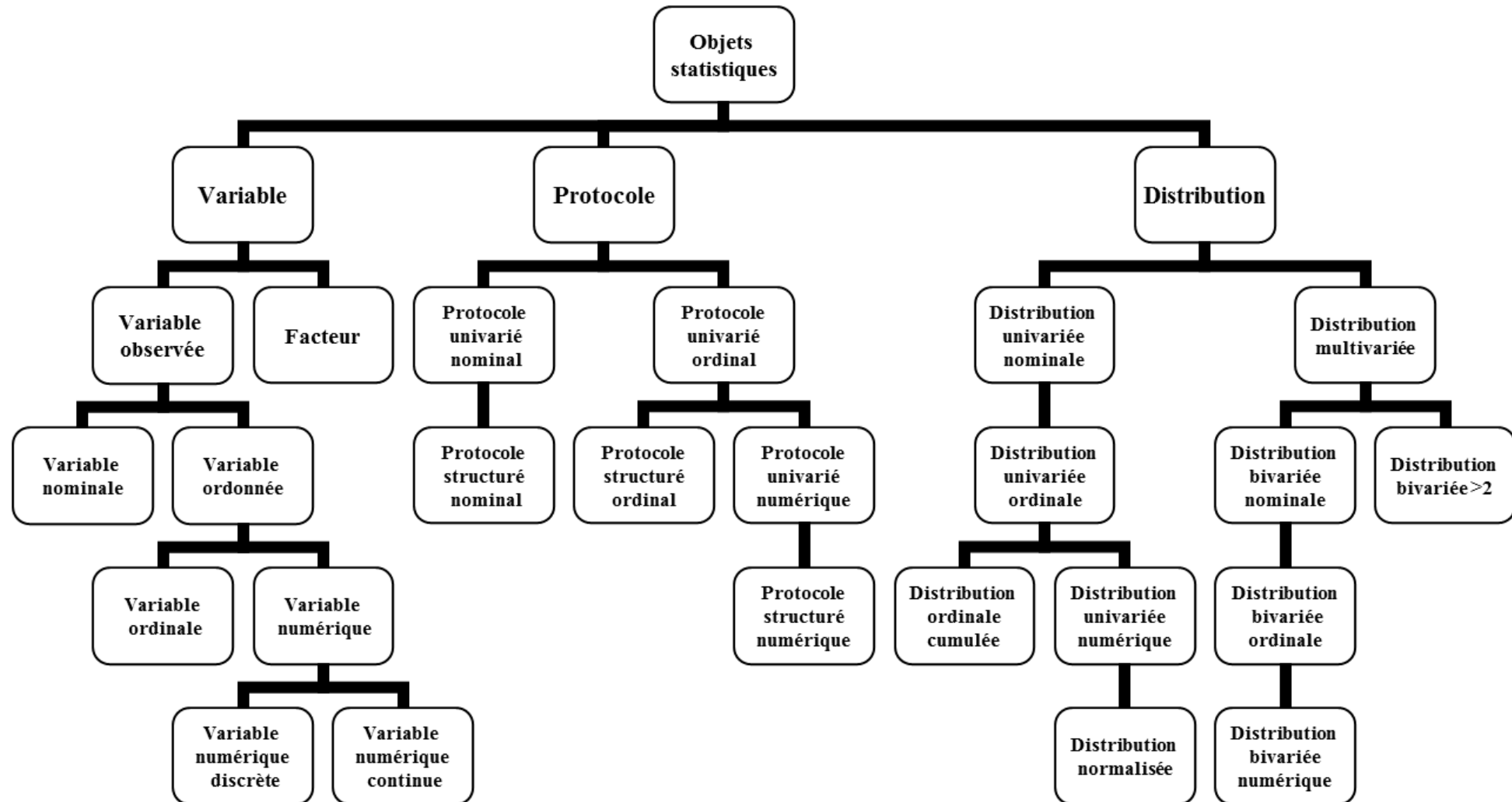
MINISTÈRE  
DE L'ÉDUCATION NATIONALE,  
DE L'ENSEIGNEMENT SUPÉRIEUR  
ET DE LA RECHERCHE

# Principes de construction (Sander & al, 2004)

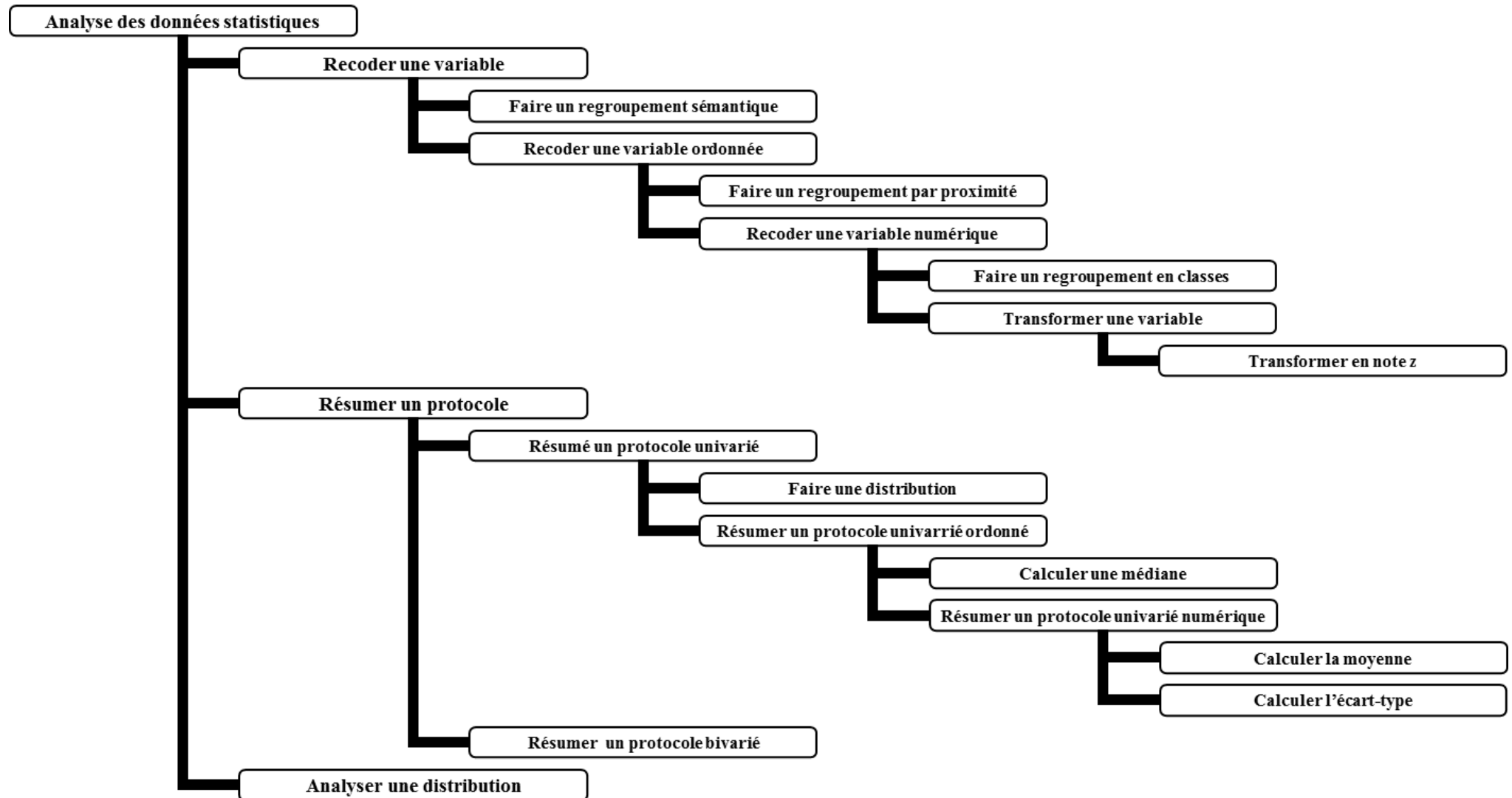
		Variable	Variable dépendante	Variable indépendante	Variable nominale	Variable ordinale	Variable numérique
Propriétés structurelles	Ensemble de modalités	X	X	X	X	X	X
	Contrôlée			X			
	Observée		X				
	Ordonnée					X	X
Propriétés fonctionnelles	Recoder	X	X		X	X	X
	Faire un regroupement sémantique				X		
	Faire un regroupement par proximité					X	
	Faire un regroupement en classe						X
	Transformer						X
	Transformer en notes Z						X



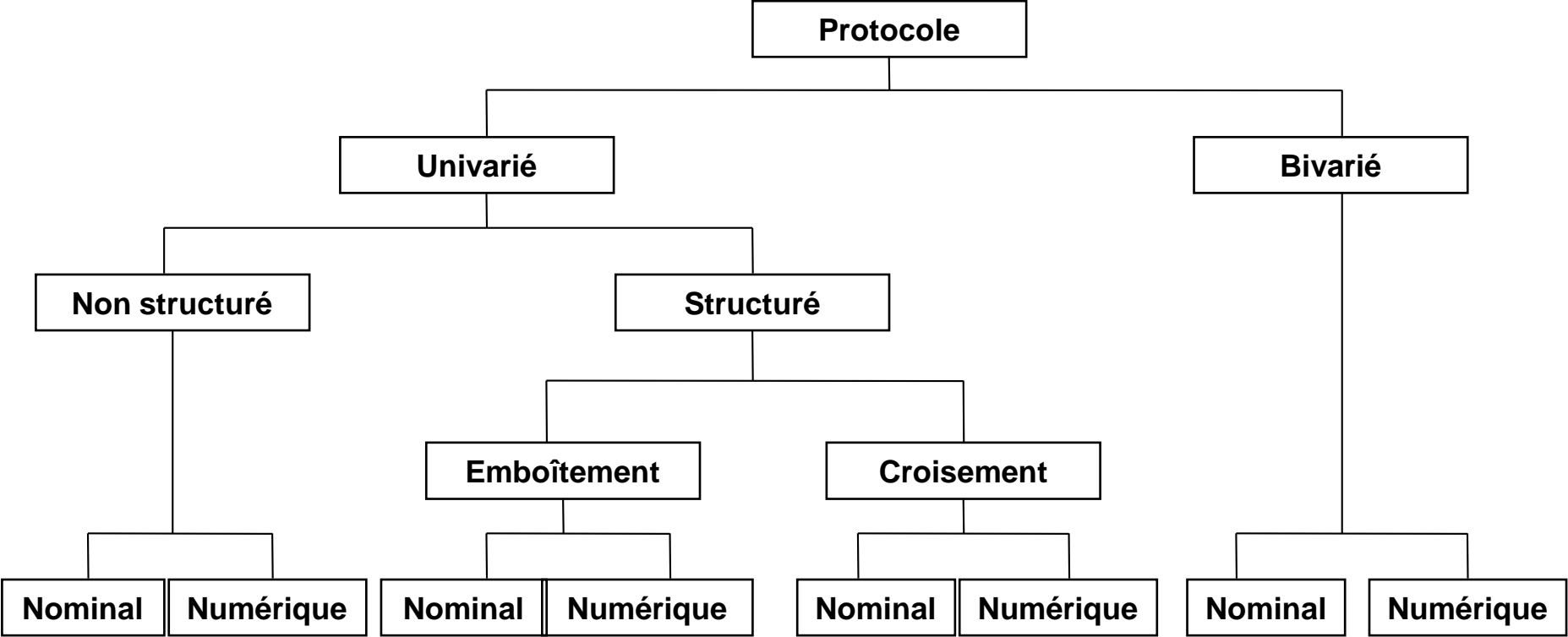
# Le réseau des concepts



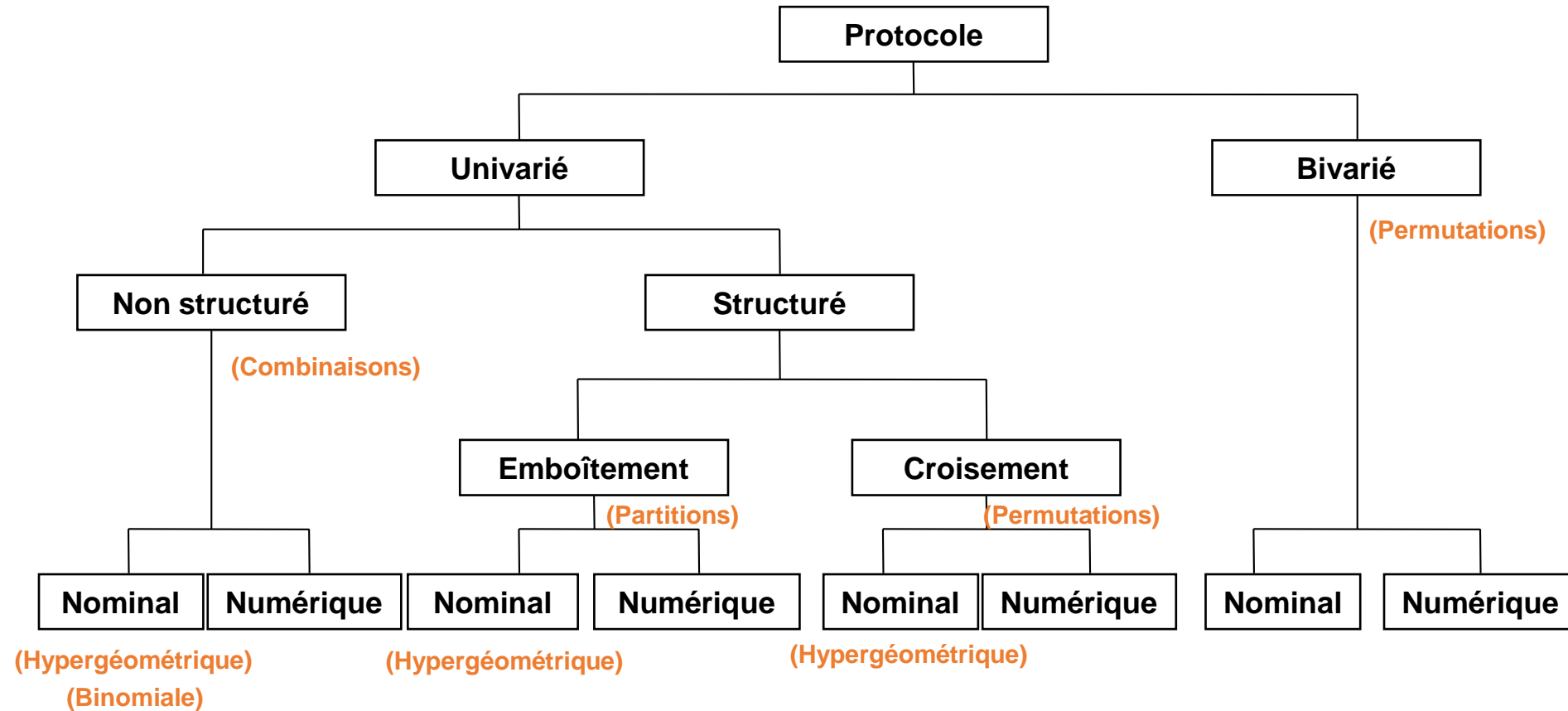
# Le réseau des procédures



# Types d'échantillons



# Distributions exactes





# Limites des réseaux sémantiques

- Un arbre de décision ?
  - How to choose a statistical test (Shankar et Singh, 2014; Marusteri & Bacarea, 2010; Yuen, 2013)
  - Autres relations entre classes que l'inclusion
- Polysémie et spécificité des termes
  - Langage quotidien vs langage scientifique
  - Variabilité des définitions; Extension, contextualisation
- Des classes peuvent être des prérequis ou des résultats pour d'autres classes
  - (ex moyenne et écart-type pour le calcul de Z; quartile pour l'écart interquartile)
- Variabilité des modes de réalisation
  - (à la main, Excel, R, SPSS etc. )

# Définition de l'ontologie

**Une ontologie est la définition formelle de la description d'un domaine de connaissance à l'aide d'un ensemble de concepts hiérarchisés**

- Les types d'ontologie(Charlet, Bachimont & Troncy , 2013)
  - l'ontologie du domaine organise les concepts
  - l'ontologie de méthodes (ex. diagnostic médical),
  - l'ontologie d'application qui articule d'une ontologie du domaine et d'une ontologie de méthode,
  - l'ontologie de représentation qui repère et organise les primitives de la théorie

# Le web sémantique

- Une ontologie est une collection de documents permettant de décrire les concepts d'un domaine à l'aide d'une taxonomie de classes et un ensemble de règle d'inférences.
- Les informations sont définies sur le web par un identifiant (URI)
- Forme et fond sont séparés (langage XML)
- Les sens est codé sous forme de triplets orientés : sujet – prédicat - objet (Langage RDF)

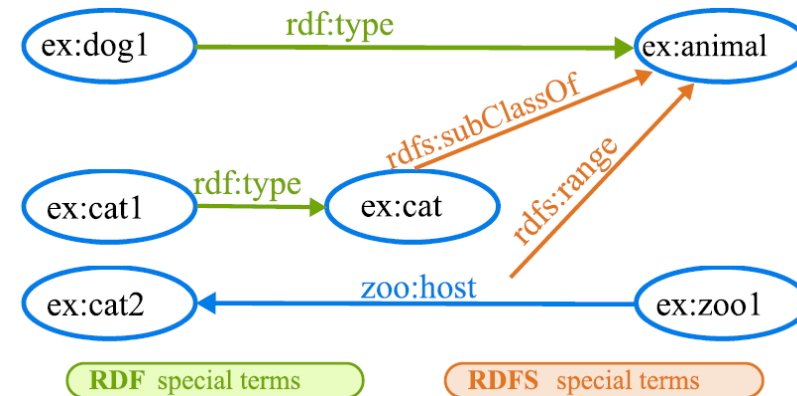
Un chien est un animal

le chat1 est un chat

Les chats sont des animaux

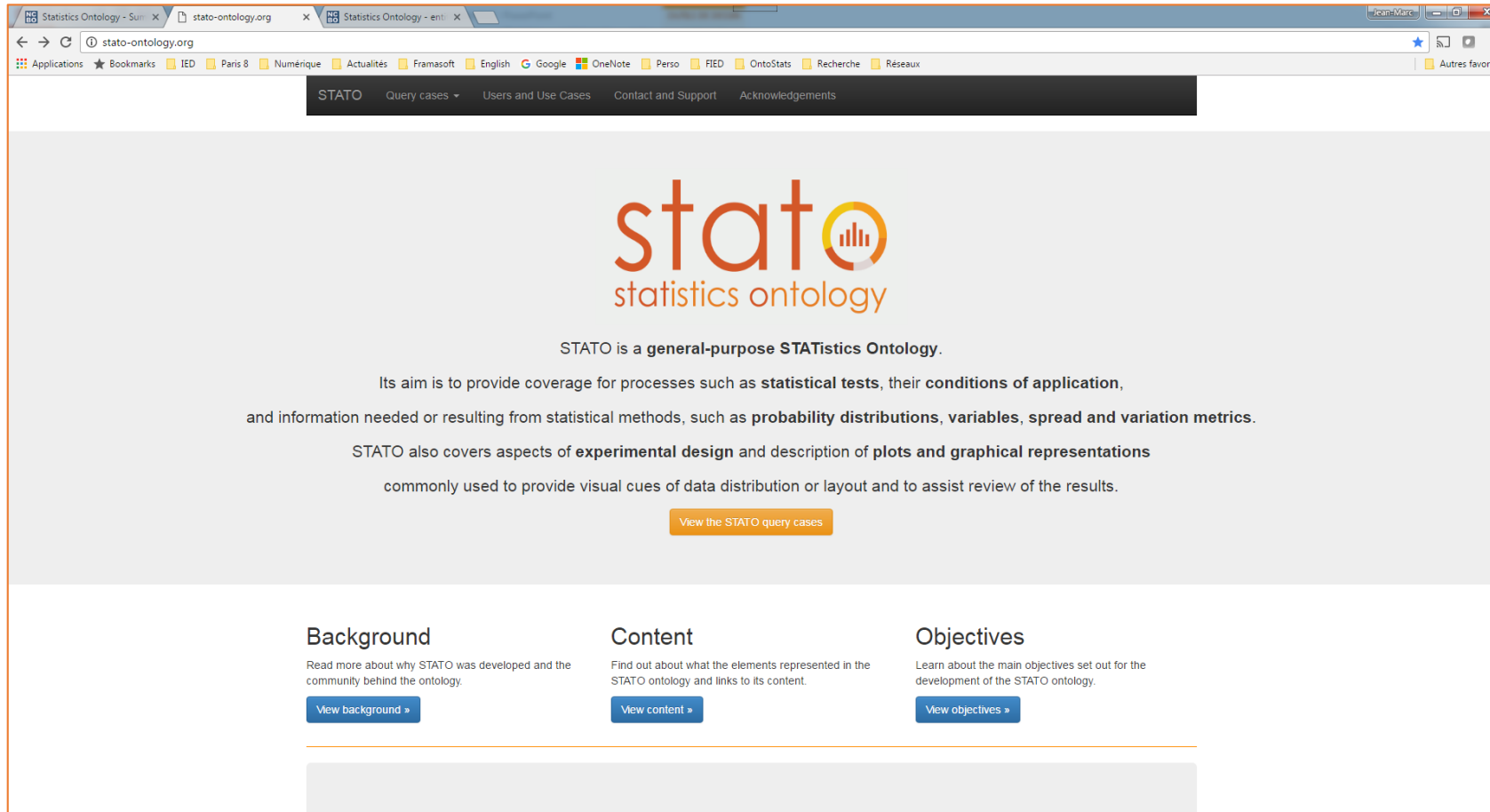
les zoos hébergent uniquement des animaux

Le zoo1 héberge le chat2





# STATO (Rocca-Serra; Gonzalez-Bertran, Burke et Sansone, 2012)



The screenshot shows the homepage of the STATO website. The browser address bar displays 'stato-ontology.org'. A navigation menu at the top includes 'STATO', 'Query cases', 'Users and Use Cases', 'Contact and Support', and 'Acknowledgements'. The main content area features the 'stato statistics ontology' logo, where 'stato' is in orange and 'statistics ontology' is in a lighter orange. Below the logo, the text reads: 'STATO is a **general-purpose STATistics Ontology**. Its aim is to provide coverage for processes such as **statistical tests**, their **conditions of application**, and information needed or resulting from statistical methods, such as **probability distributions**, **variables**, **spread and variation metrics**. STATO also covers aspects of **experimental design** and description of **plots and graphical representations** commonly used to provide visual cues of data distribution or layout and to assist review of the results.' A yellow button labeled 'View the STATO query cases' is positioned below this text. At the bottom, there are three columns: 'Background' with a 'View background »' button, 'Content' with a 'View content »' button, and 'Objectives' with a 'View objectives »' button.

STATO is a **general-purpose STATistics Ontology**.

Its aim is to provide coverage for processes such as **statistical tests**, their **conditions of application**, and information needed or resulting from statistical methods, such as **probability distributions**, **variables**, **spread and variation metrics**.

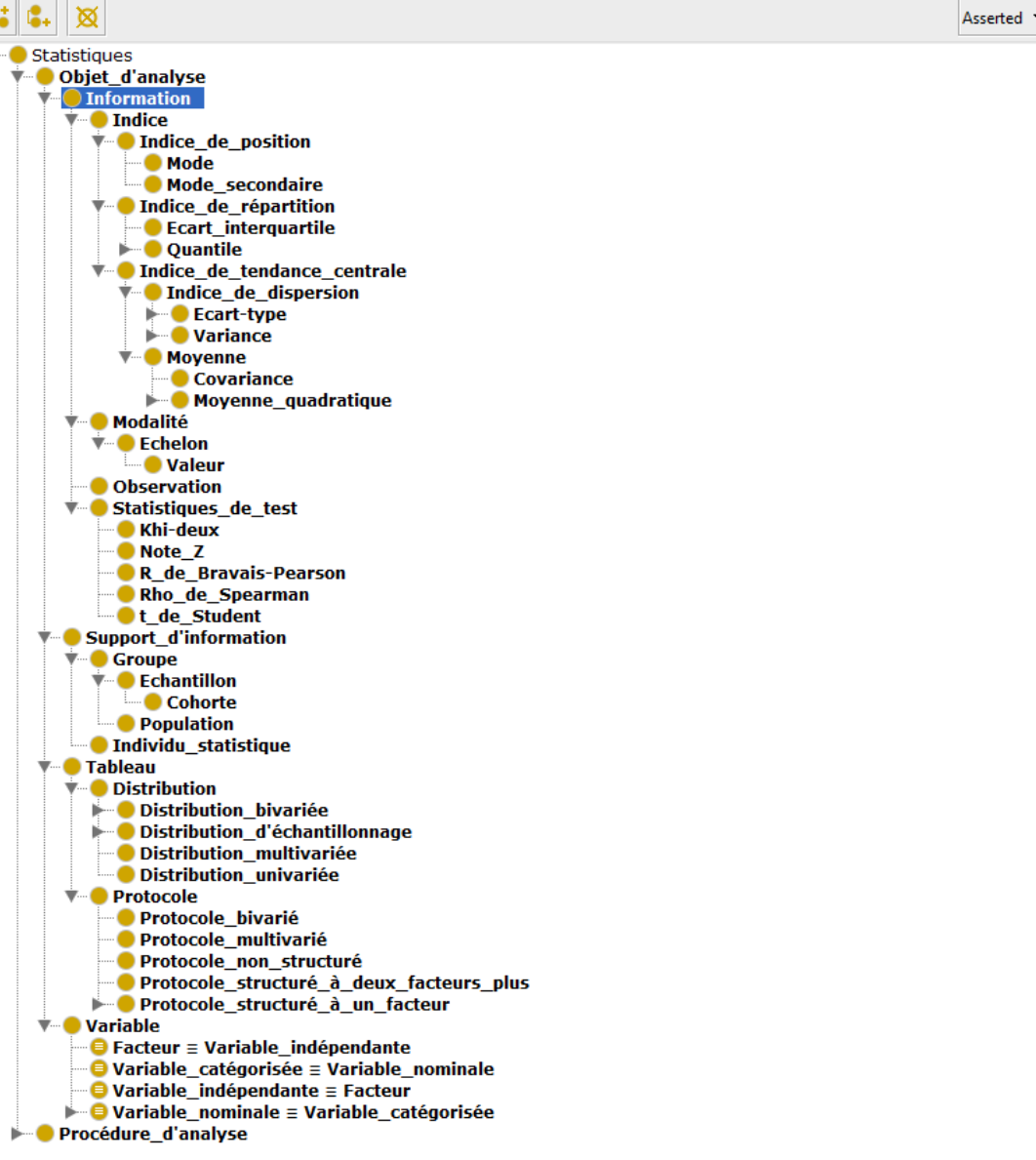
STATO also covers aspects of **experimental design** and description of **plots and graphical representations** commonly used to provide visual cues of data distribution or layout and to assist review of the results.

[View the STATO query cases](#)

**Background**  
Read more about why STATO was developed and the community behind the ontology.  
[View background »](#)

**Content**  
Find out about what the elements represented in the STATO ontology and links to its content.  
[View content »](#)

**Objectives**  
Learn about the main objectives set out for the development of the STATO ontology.  
[View objectives »](#)



Annotations +

`rdfs:comment` [language: fr]  
Ensemble de valeurs manipulées dans les procédures statistiques

Equivalent To +

SubClass Of +  
● **Objet\_d'analyse**

General class axioms +

SubClass Of (Anonymous Ancestor)

Instances +

Target for Key +

Disjoint With +

Disjoint Union Of +



Asserted



Annotations +

Description: Procédure\_d'analyse

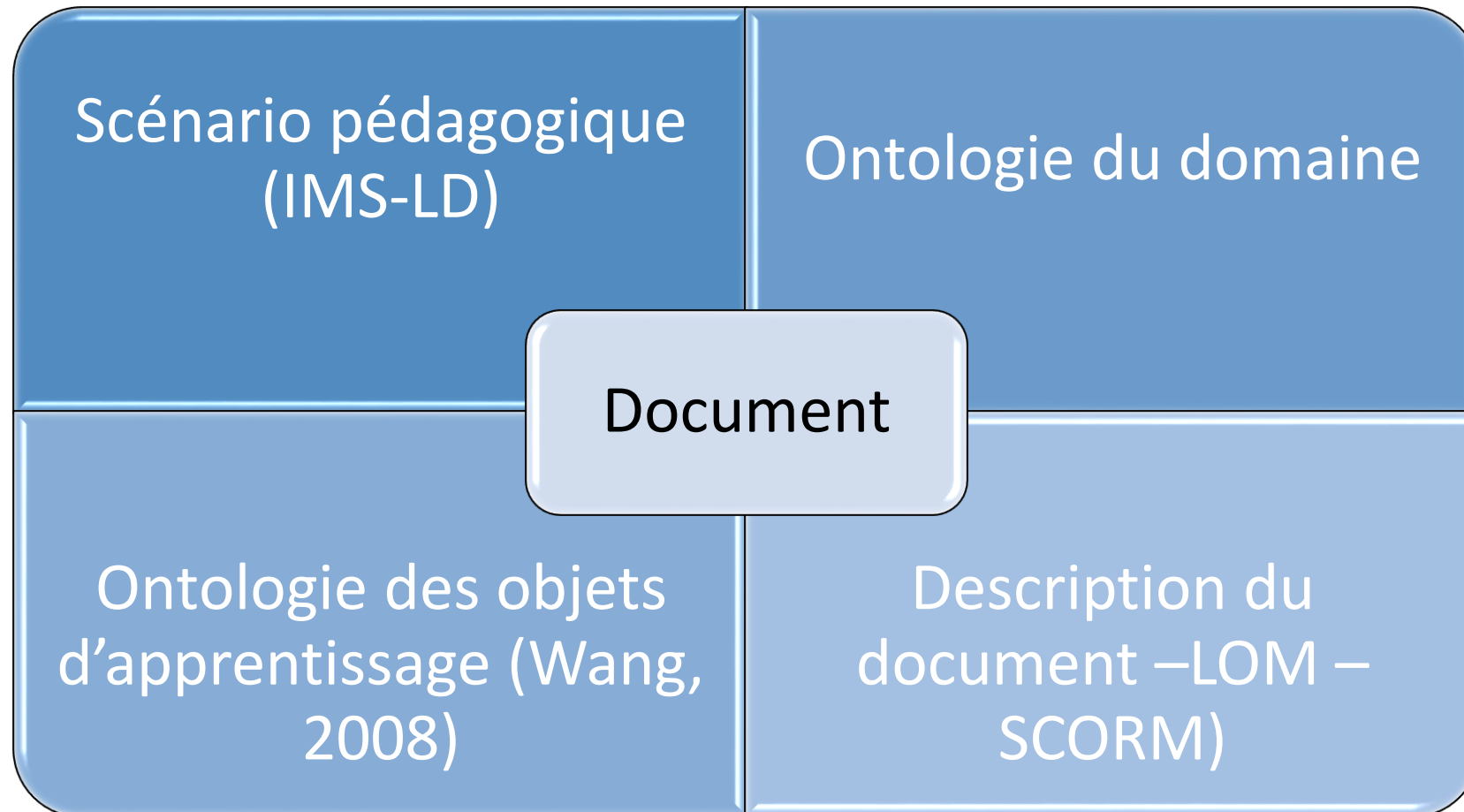
- Equivalent To +
- SubClass Of +
- General class axioms +
- SubClass Of (Anonymous Ancestor)
- Instances +
- Target for Key +
- Disjoint With +
- Disjoint Union Of +

# Perspectives d'application

# L'indexation des REL (Oufia, Abel et Moulin, 2005)

- Learning object metadata (LOM) : Normes de description des ressources pédagogiques
  - Processus d'indexation long et fastidieux (Michel 2004)
  - Problèmes d'ambiguïté de certains éléments du LOM (Najjar & Duval, 2006)
  - Problèmes de compréhension chez les créateurs de ressources (Currier, 2004)
- Décrire la structure, le contenu et les stratégies d'apprentissage d'un objet pédagogique avec des ontologies. Mohan & Brooks, 2003; Hernandez, Mothe, Ralalason et Stolf, 2006)

# Les multiples facettes des REL (Hernandez, Mothe, Ralalason et Stolf, 2006)



# Mobilisation de la communauté



## Enseignant de statistiques

- Validation de l'ontologie
- Partage des ressources
- Diversification des pratiques

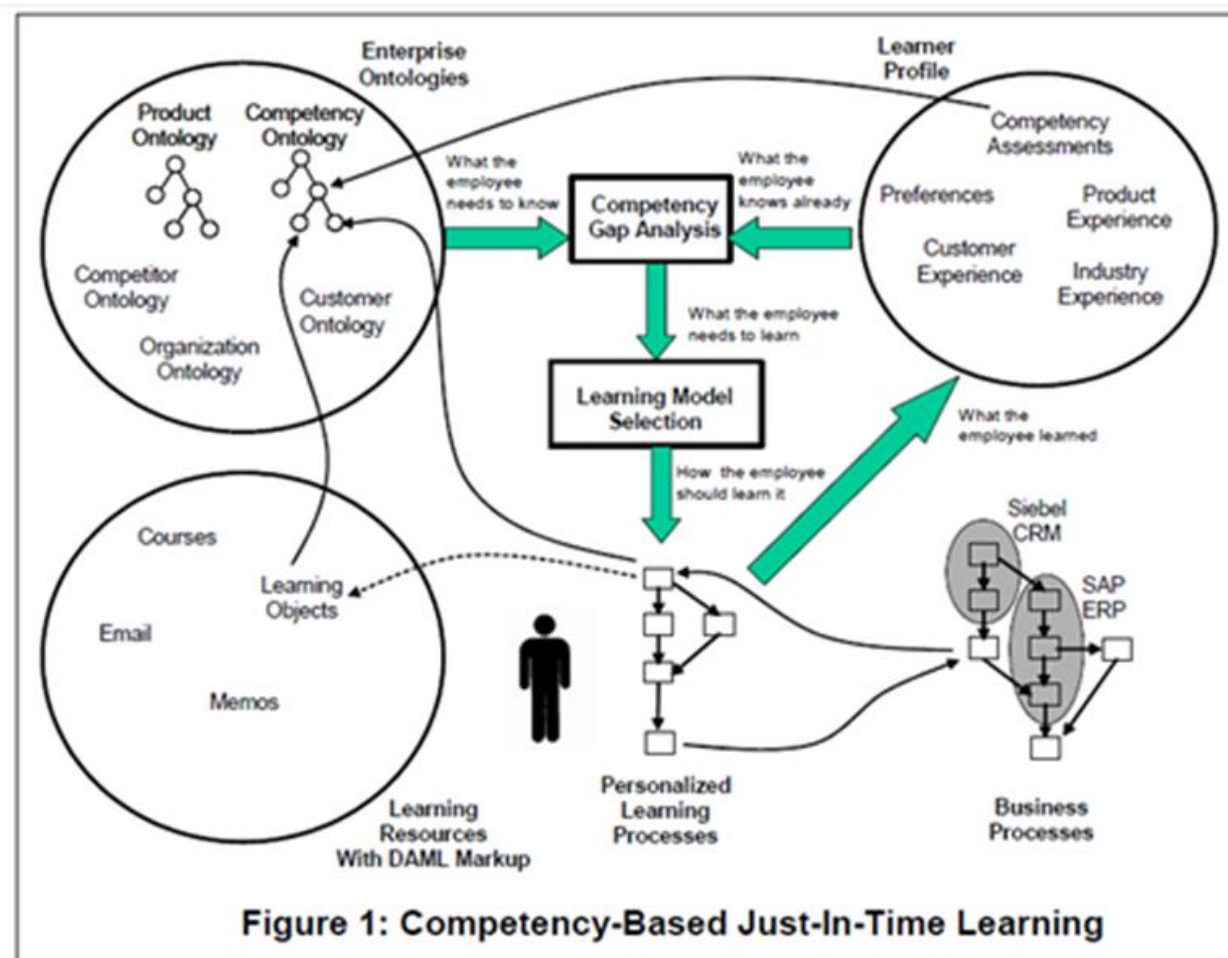
## UNT et ontologues

- Enrichissement de l'ontologie
- Moteur d'inférence
- Aide à l'indexation

## Psychologues et sciences de l'éducation

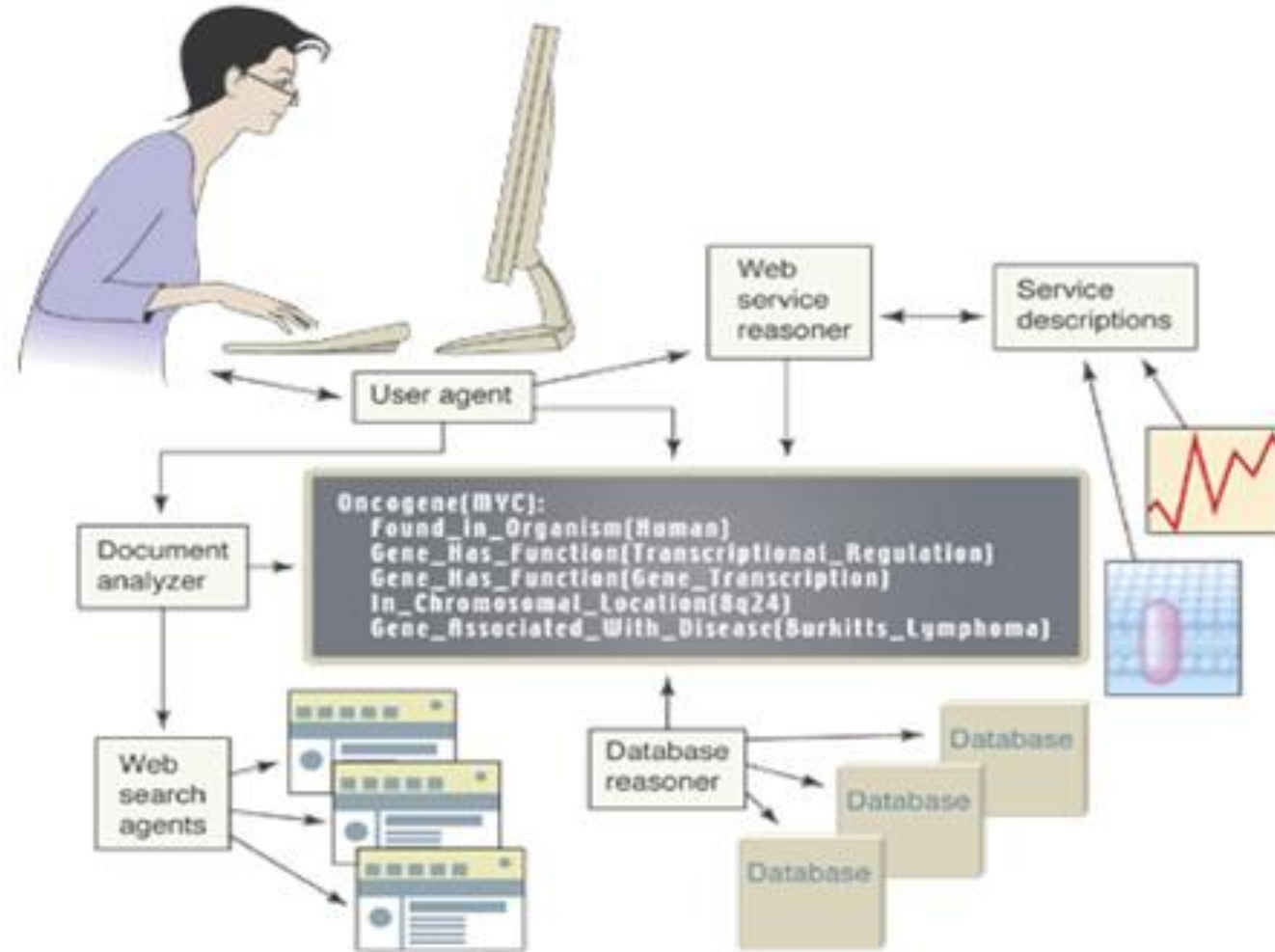
- Etude de l'apprentissage
- Usages des ressources éducatives

# Environnement d'apprentissage personnalisé





# Environnement d'apprentissage adaptatif



Merci pour votre attention