

---

# Instrumenter l'interprétation dans une démarche centrée sur l'utilisateur

---

Pierre Beust

GREYC CNRS UMR 6072 & Pôle ModesCoS de la MRSH  
Université de Caen - Basse Normandie  
14032 Caen Cedex, France  
[pierre.beust@unicaen.fr](mailto:pierre.beust@unicaen.fr)



# Plan

---

1. Mon parcours
2. Objectifs et problématique
3. Travaux et développements
4. Expérimentations sur corpus
5. Sémantique Interprétative et TAL
6. Vers des ENT énatifs
7. Conclusions

# 1. Mon parcours

---

- Qui suis-je ?
  - Maître de conférence en Informatique
  - Laboratoire GREYC CNRS UMR 6072
  - Université de Caen Basse-Normandie
  - Directeur d'un service commun FOAD/TICE et responsable C2I-1
  
- Avant ...
  - DEA IA (1992)
  - Vacataire Université Antilles-Guyane (1992-1995)
  - Thèse sous la co-dir. de A. Nicolle et L. Gosselin (1995-1998)

# Plan

---

1. Mon parcours
- 2. Objectifs et problématique**
3. Travaux et développements
4. Expérimentations sur corpus
5. Sémantique Interprétative et TAL
6. Vers des ENT énatifs
7. Conclusions

## 2. Objectifs et problématique

---

Des centres d'intérêt :

- Informatique & Linguistique

  - ↳ Traitement Automatique des Langues

- Sémantique des textes

  - ↳ Linguistique de corpus, Sémantique Interprétative

- Interactions Homme-Machine centrées sur l'utilisateur

- Approches pluridisciplinaires de la complexité et des sciences cognitives

## 2. Objectifs et problématique

---

- Des « repères » pour penser (quelques interprétations personnelles d'auteurs) :

- La signification se justifie par des rapports de similarité et de différence

*CLG, F de Saussure*

- L'interprétation est une activité silencieuse mais c'est une activité langagière tout aussi fondamentale que celle du locuteur

*La fonction muette du langage, J. Coursil*

- Le sens est une perception sémantique située, dont toute tentative d'objectivation (et donc de représentation) est une sommation incomplète de points de vue.

*Sémantique Interprétative, F. Rastier*

## 2. Objectifs et problématique

---

- Ancrage épistémologique :
  - Sémantique Interprétative (François Rastier) :
    - perception sémantique
    - dynamique sémique (notions d'isotopie, d'afférence, d'opération interprétative)
    - principe de la détermination du local par le global (alternative à la compositionnalité)
  - Approche interactionniste et expérimentale :
    - des logiciels d'étude pour monter des protocoles expérimentaux
    - expérimenter des conditions d'interprétation
    - tracer les interprétations
    - des interactions, des traitements et ressources avant tout centrés sur l'utilisateur
    - l'étude des usages dans les ENT (usages, contournements, détournements ...)

## 2. Objectifs et problématique

---

- Objectifs :
  - Contribuer au développement d'Environnements Numériques de Travail intégrant des aspects sémiotiques
    - Veille documentaire, Analyse de flux d'information  
Recherche d'information, ...
- Propositions :
  - Les ressources lexicales propres à l'utilisateur
  - Subjectiver les traitements
  - La cartographie thématique personnalisée
  - Une assistance à l'interprétation, à l'accès personnalisé aux contenus dans des collections documentaires

# Plan

---

1. Mon parcours
2. Objectifs et problématique
- 3. Travaux et développements**
4. Expérimentations sur corpus
5. Sémantique Interprétative et TAL
6. Vers des ENT énatifs
7. Conclusions

# 3. Travaux et développements

---

- Depuis une quinzaine d'années, plusieurs projets de développement et projets de recherche
  - Avec des étudiants
  - Avec des doctorants (V. Perlerin, T. Roy)
  - Avec des collègues universitaires
    - Groupe NU (Nouveaux Usages)
  - Avec des partenaires industriels (ExoMakina)

# 3. Travaux et développements

---

- Plusieurs réalisations logicielles :
  - Anadia (1995-1999)
    - Représentation lexicale et analyse par calcul d'isotopies
  - ThemeEditor (2000-)
    - Coloriage d'isotopies génériques
  - Lucia (2000-)
    - Structuration du contenu lexical
    - Application à la recherche d'information
  - ProxiDocs (2004-)
    - Cartographie thématique de corpus
  - CartoMail (2009-2010)
    - Cartographie de boites mails
  - Canopée (2011)
    - Assistance interprétative pour les collections documentaires

# 3. Travaux et développements

---

- Anadia
  - Une méthodologie de Jacques Coursil
    - Attributs-Tables-Topiques
      - Ex : les phonèmes du français
  - Utilisation pour de la représentation lexicale
    - Structuration du contenu
    - Modèle oppositionnel du sème
  - Recherche d'isotopie
  - Application à l'analyse d'énoncés de dialogues

# 3. Travaux et développements

- Anadia

The screenshot displays a software interface with three overlapping windows:

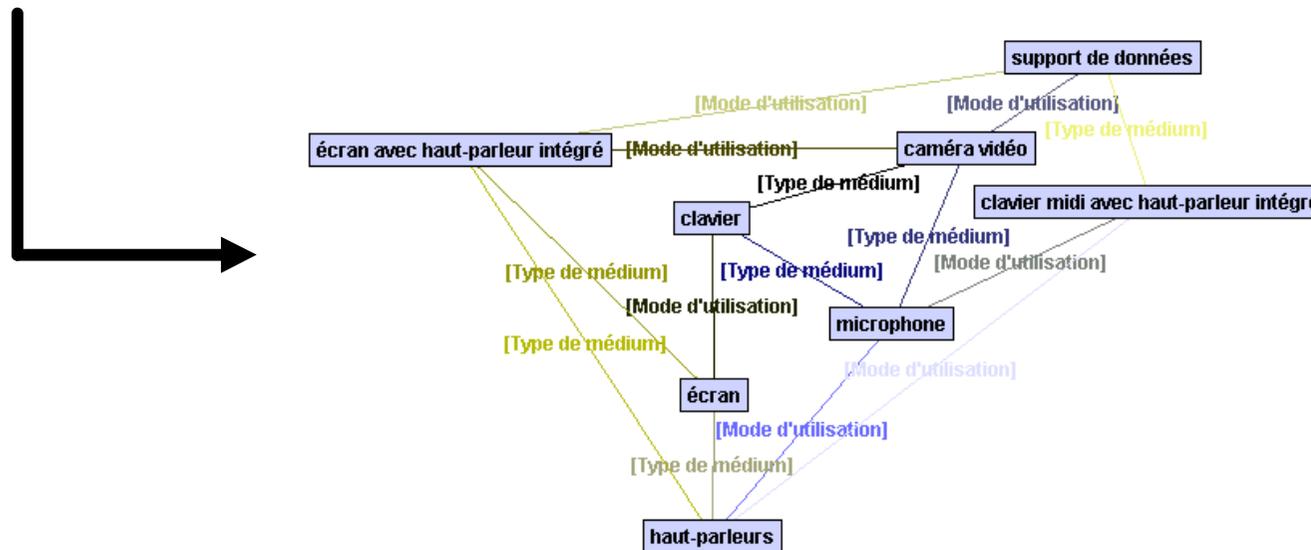
- Nouvel attribut ... (left):** Shows the definition of the 'propulsion' attribute with values 'avec moteur' and 'sans moteur'. An 'OK' button is visible.
- Nouvel attribut ... (middle):** Shows the definition of the 'Milieu' attribute with values 'air' and 'terre'.
- Table : TRANSPORT-TABLE:** A table with columns 'MILIEU' and 'PROPULSION'.

|   | MILIEU | PROPULSION  |
|---|--------|-------------|
| 6 |        |             |
| + | AIR    | AVEC-MOTEUR |
| + | AIR    | SANS-MOTEUR |
| + | TERRE  |             |
| + | TERRE  |             |
| + | MER    |             |
| + | MER    |             |
- Topique en graphe : TRANSPORT-TOPIQUE75:** A graph showing relationships between transport modes: 'avion', 'planeur', 'automobile', 'cycle', 'bateau à moteur', and 'voilier'.

# 3. Travaux et développements

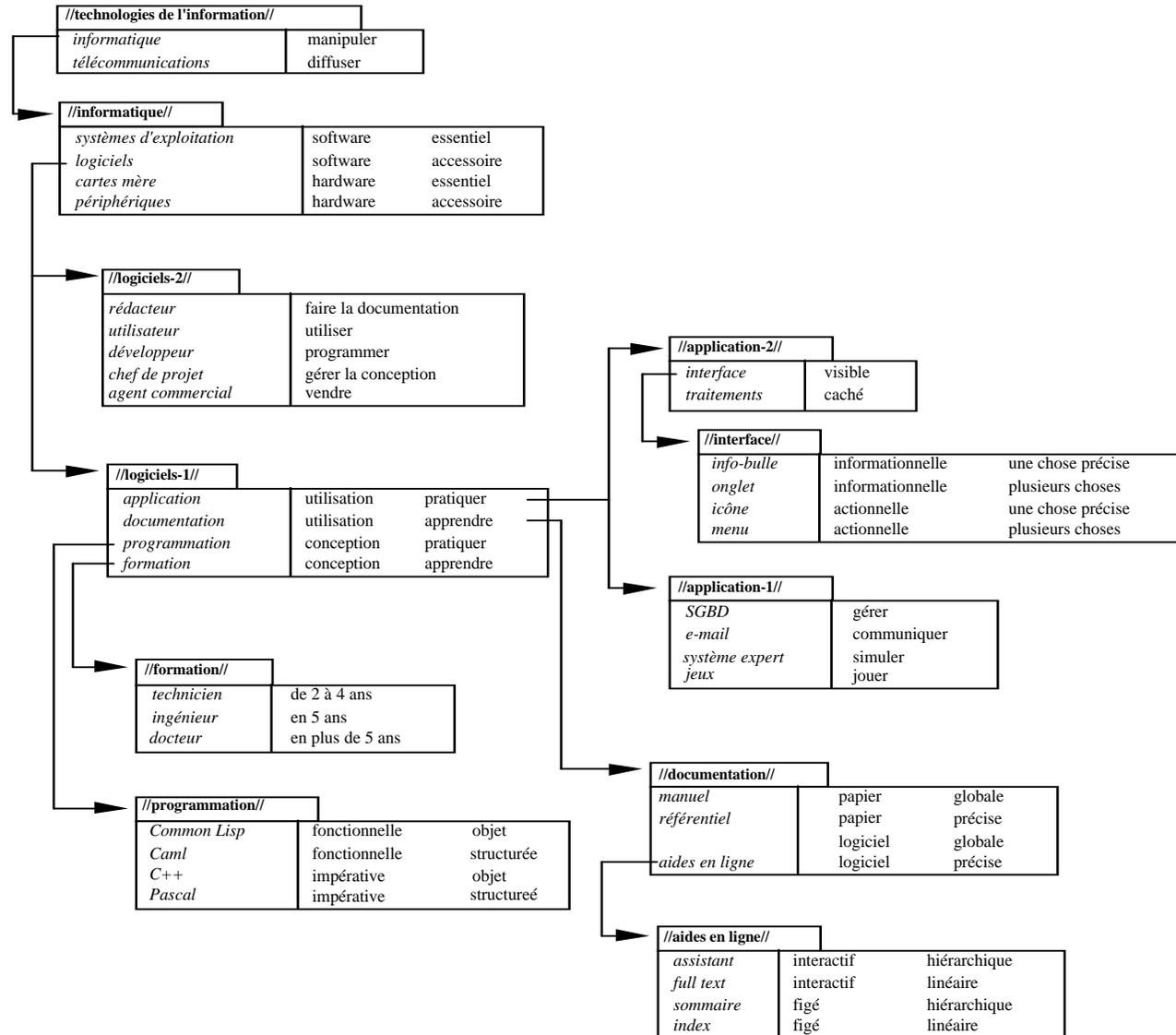
- Anadia

|   | Type de médium        | Mode d'utilisation |
|---|-----------------------|--------------------|
| clavier, souris, appareil photo numérique   | image et texte        | entrée             |
| caméra vidéo  | image et texte et son | entrée             |
| microphone, micro   | son                   | entrée             |
| écran, imprimante   | image et texte        | sortie             |
| écran avec haut-parleur intégré, écran avec hauts-parleurs intégrés               | image et texte et son | sortie             |
| haut-parleurs, HP, enceintes, enceinte, haut-parleur                              | son                   | sortie             |
|   | image et texte        | entrée / sortie    |
| support de données  | image et texte et son | entrée / sortie    |
| clavier midi avec haut-parleur intégré, clavier midi avec hauts-parleurs intégrés | son                   | entrée / sortie    |



# 3. Travaux et développements

- Anadia



# 3. Travaux et développements

- Anadia : représentation du contenu lexical

| Utilisation de l'information | Type de composant               | Rôle du composant              | Rapport au logiciel              | Pratique du logiciel          | Présentation de l'information | Support de l'information  | Accès à l'information     | Structure de l'information      |
|------------------------------|---------------------------------|--------------------------------|----------------------------------|-------------------------------|-------------------------------|---------------------------|---------------------------|---------------------------------|
| diffuser<br><b>manipuler</b> | matériel<br><b>non matériel</b> | essentiel<br><b>accessoire</b> | conception<br><b>utilisation</b> | pratiquer<br><b>apprendre</b> | globale<br><b>précise</b>     | papier<br><b>logiciel</b> | <b>interactif</b><br>figé | <b>hiérarchique</b><br>linéaire |

*Classème* *Sémantème*

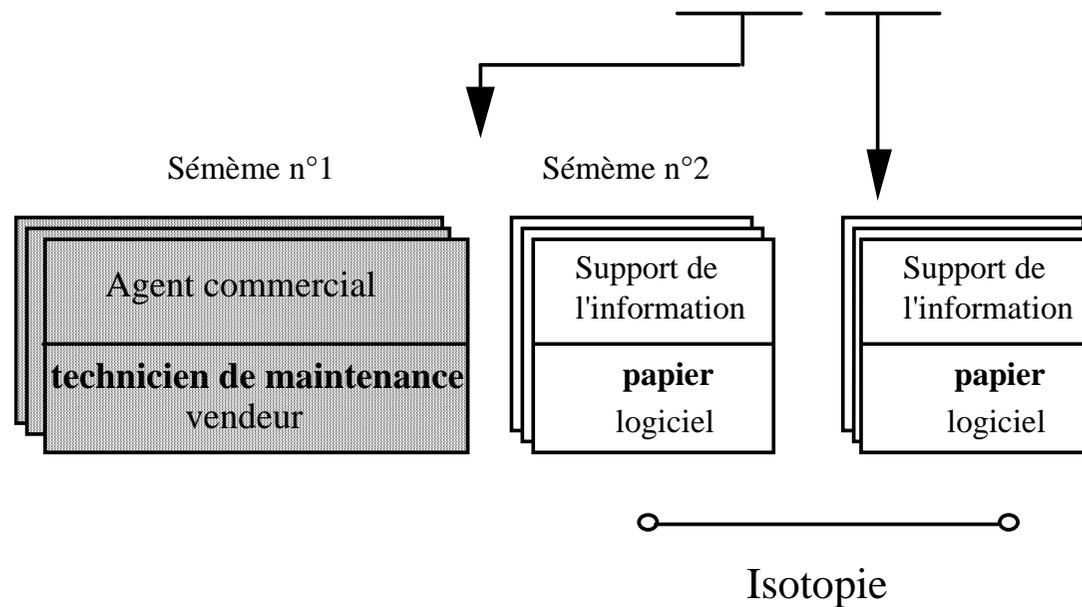
La lexie « assistant »

### 3. Travaux et développements

---

- Anadia : recherche d'isotopies dans des énoncés de dialogue

*... je me débrouille toute seule avec mon support papier*



# 3. Travaux et développements

---

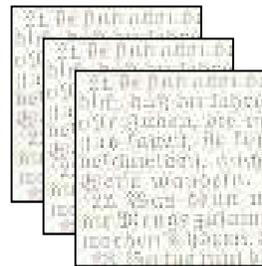
- ThemeEditor
  - Visualiser sur corpus des isotopies génériques
  - Utiliser une représentation lexicale plus intuitive
    - « Sacs de mots » plus qu'une représentation de contenu
  - Mettre en place une interaction ressources-corpus

# 3. Travaux et développements

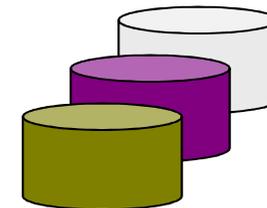
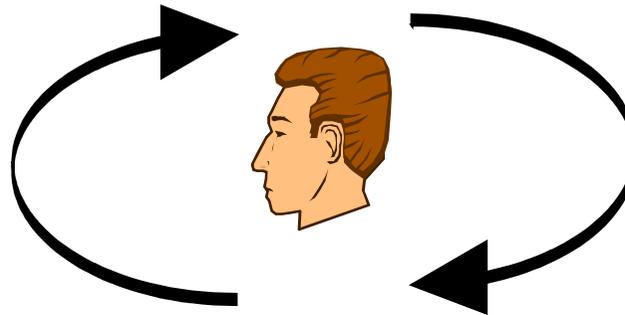
- ThemeEditor

```
mot1  n1  
mot2  n2  
...
```

Liste d'occurrences



Corpus de textes électroniques  
(corpus homogène)



Classes sémantiques

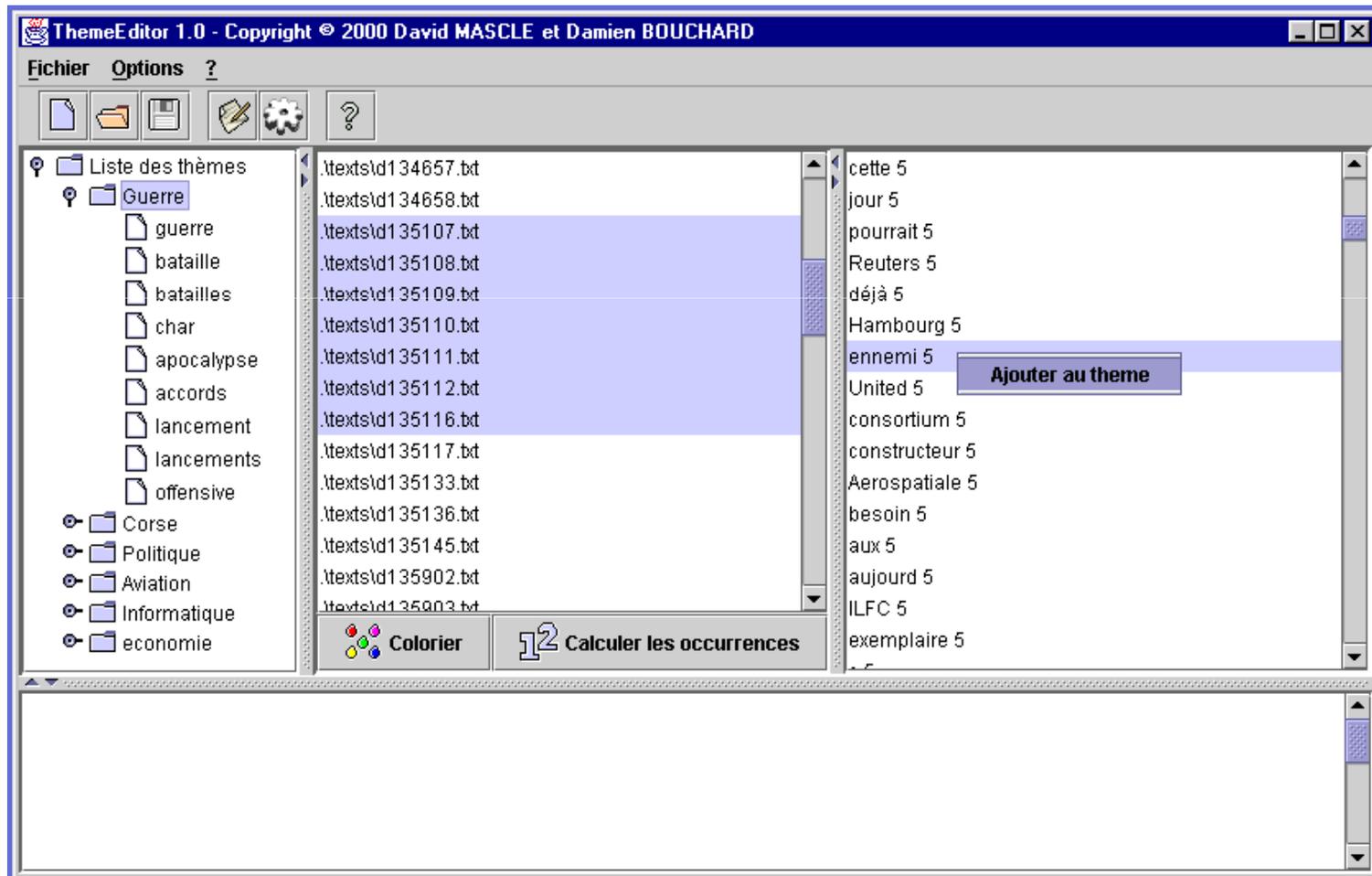
| Classe | Couleur    | Nb. mots | % total |
|--------|------------|----------|---------|
|        | Vert olive |          |         |
|        | Jaune      |          |         |
|        | Blanc      |          |         |

Statistiques de coloriage

↳ représentations de Thèmes

# 3. Travaux et développements

- ThemeEditor



# 3. Travaux et développements

- ThemeEditor

The screenshot shows the ThemeEditor 1.0 application window. The title bar reads "ThemeEditor 1.0 - Copyright © 2000 David MASCLE et Damien BOUCHARD". The menu bar includes "Fichier", "Options", and "?". The toolbar contains icons for file operations and settings. The main interface is divided into three sections: a left sidebar with a tree view of themes (Guerre, Corse, Politique, Aviation, Informatique, économie), a central list of text files (.textsid134657.bt to .textsid135136.bt), and a right pane displaying a list of words with their occurrence counts (e.g., acheter 3, actionnaires 3). Below the file list are buttons for "Colorier" and "Calculer les occurrences". At the bottom, a table summarizes the data for two themes.

| Classement | Définition du thème | Nombre de mots appartenant au thème | Couleur | Nombre de mots coloriés | Nombre de mots différents utilisés pour colorier | Pourcentage de recouvrement du thème | Pourcentage (par rapport au nombre de mots coloriés) | Pourcentage (par rapport au nombre de mots) |
|------------|---------------------|-------------------------------------|---------|-------------------------|--|--------------------------------------|--|---|
| 1          | Aviation            | 46                                  | Orange  | 19                      | 11   | 23,91 %                              | 86,36 %  | 9,05 %                                      |
| 2          | économie            | 99                                  | Vert    | 3                       | 1  | 1,01 %                               | 13,64 %  | 1,43 %                                      |

# 3. Travaux et développements

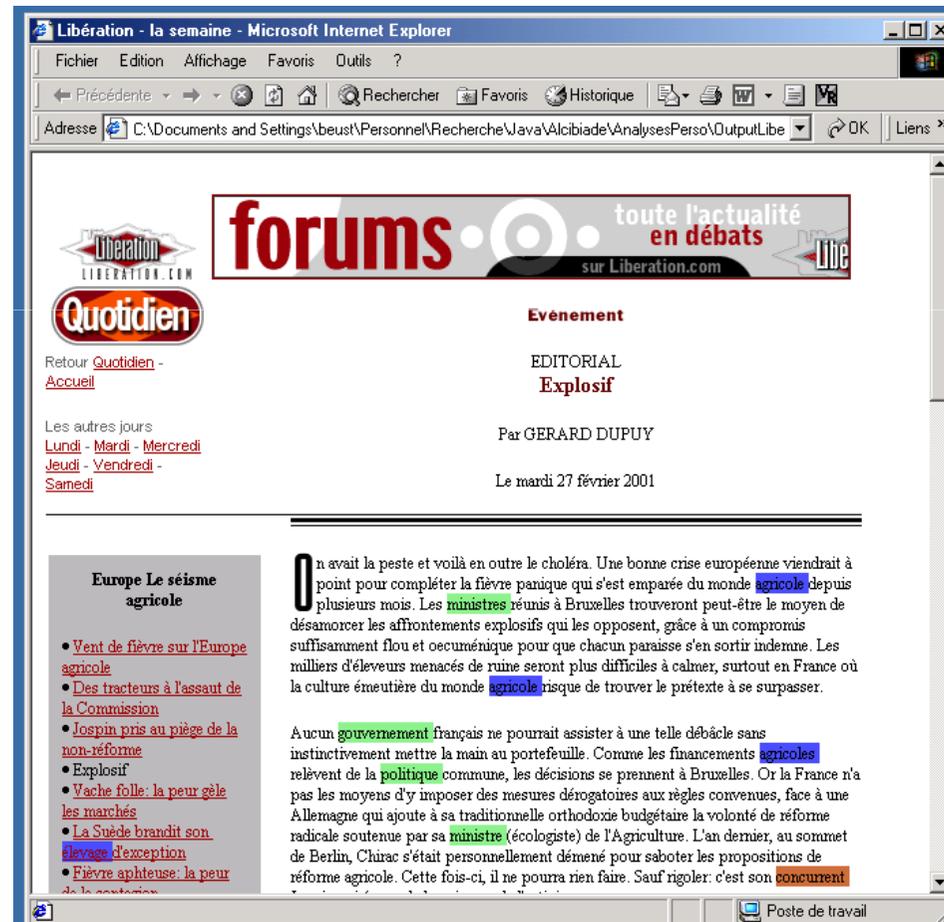
- ThemeEditor

The screenshot shows the ThemeEditor application interface. The top window displays a text document with several words highlighted in different colors (yellow, green, purple, red). The text discusses economic and financial topics, mentioning figures like Jean-Claude Trichet and Dominique Strauss-Kahn, and terms like 'économie', 'politique', and 'guerre'. The bottom window displays a table summarizing the word analysis results.

| Classement | Définition du thème | Nombre de mots appartenant au thème | Couleur | Nombre de mots coloriés | Nombre de mots différents utilisés pour colorier | Pourcentage de recouvrement du thème | Pourcentage (par rapport au nombre de mots coloriés) | Pourcentage (par rapport au nombre de mots) |
|------------|---------------------|-------------------------------------|---------|-------------------------|--|--------------------------------------|--|---|
| 1          | economie            | 99                                  | purple  | 27                      | 16   | 16,16 %                              | 72,97 %  | 3,18 %                                      |
| 2          | Politique           | 18                                  | yellow  | 7                       | 5  | 27,78 %                              | 18,92 %  | 0,82 %                                      |
| 3          | Guerre              | 9                                   | red     | 3                       | 1  | 11,11 %                              | 8,11 %   | 0,35 %                                      |

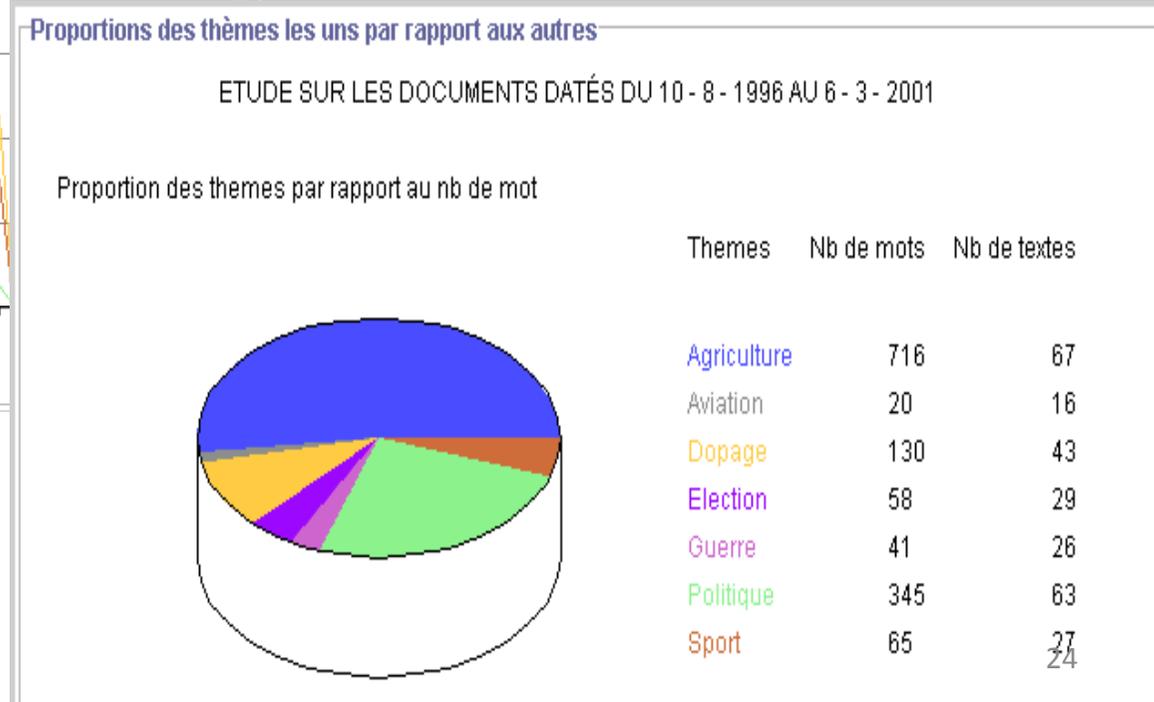
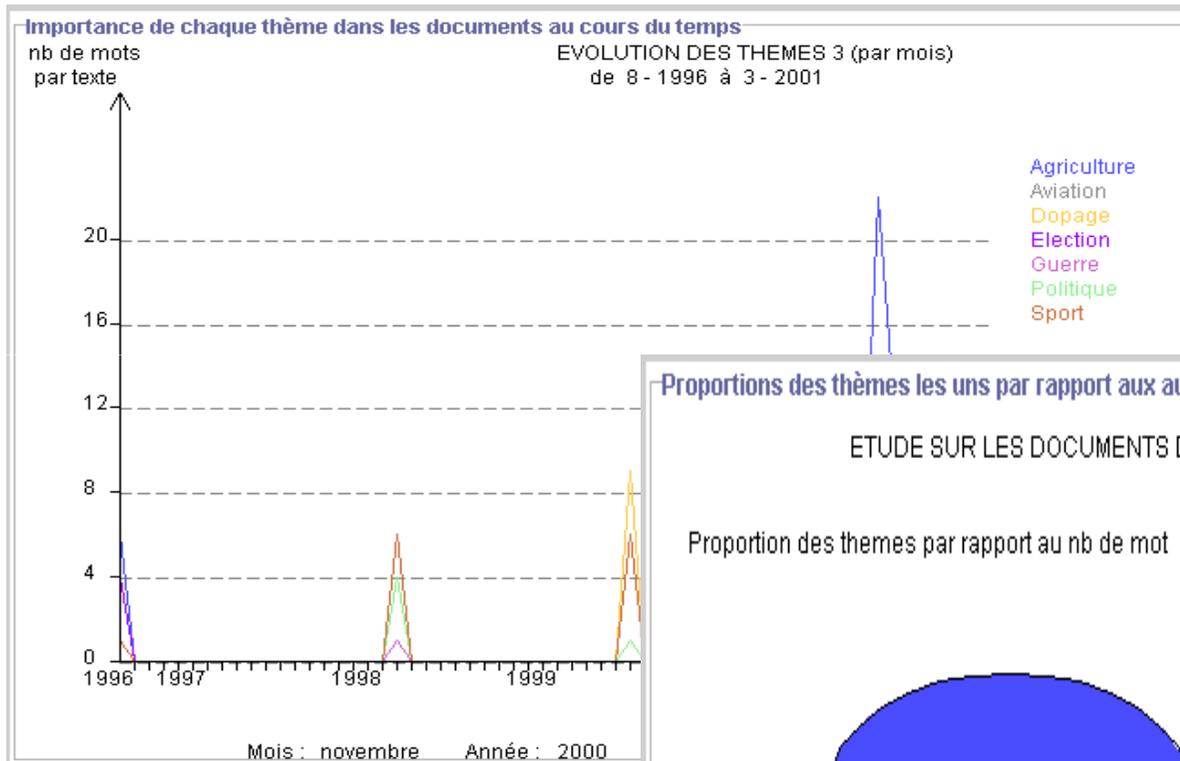
# 3. Travaux et développements

- ThemeEditor



# 3. Travaux et développements

- ThemeEditor



# 3. Travaux et développements

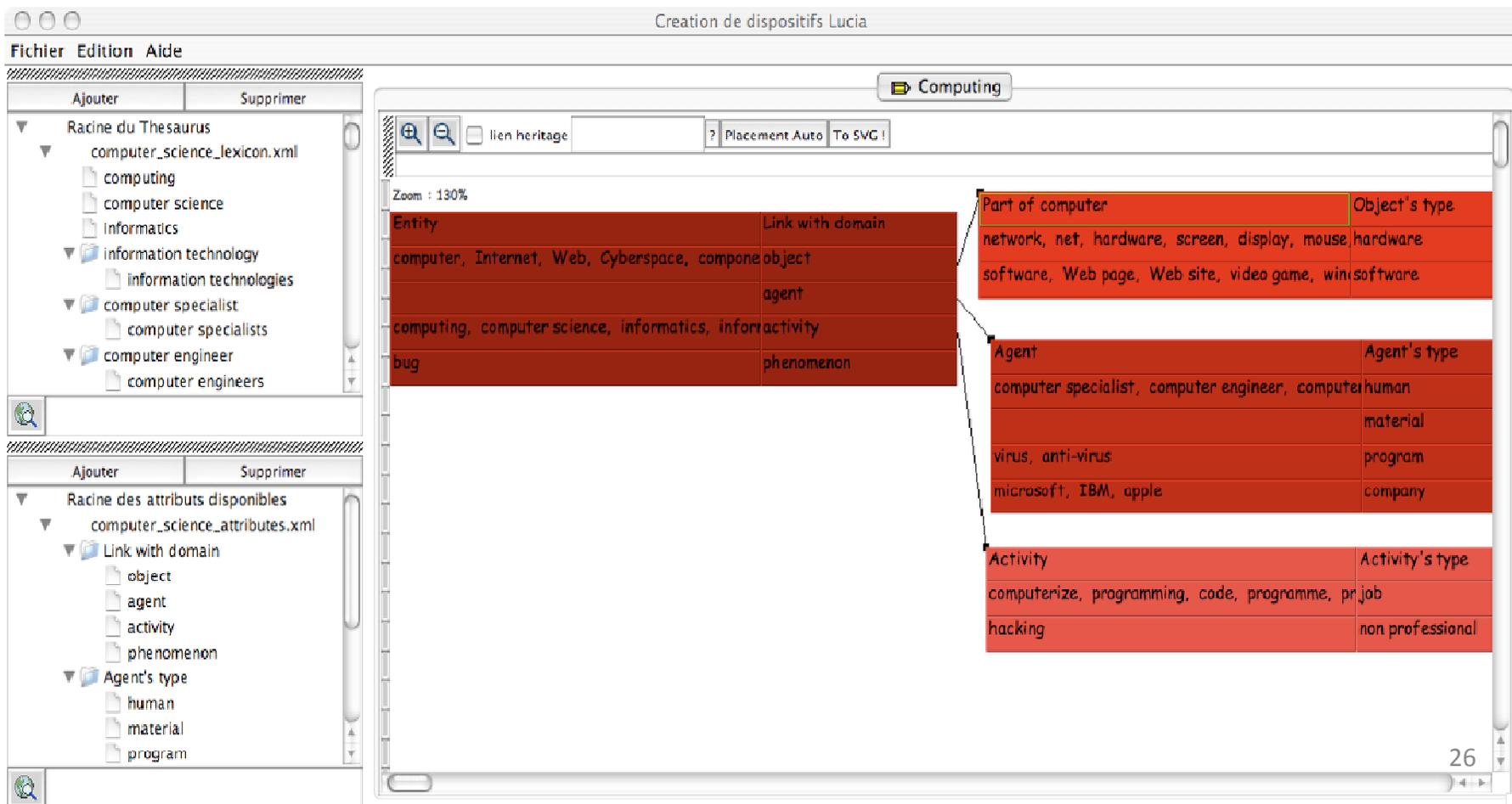
---

- LUCIA (Thèse de Vincent Perlerin)
  - Sémantique légère centrée sur l'utilisateur
    - Représentation différentielle personnelle du contenu lexical
    - Légère : représentations non exhaustives subjectivées
    - Projections rapides sur corpus
    - Applications pour la recherche documentaire sur le web

# 3. Travaux et développements

- LUCIA

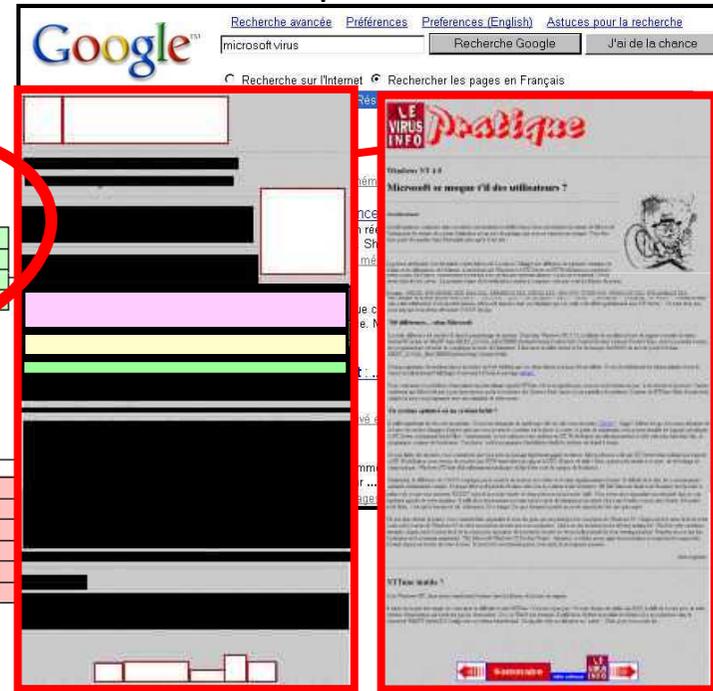
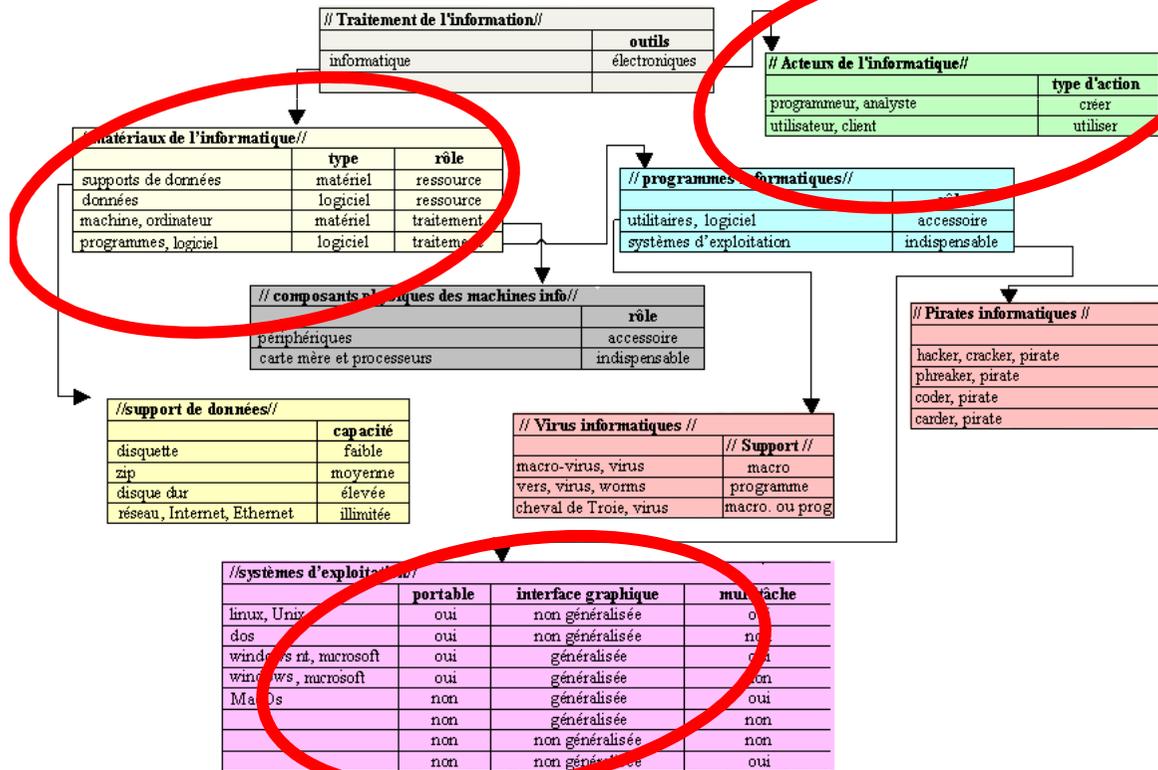
- VisualLUCIABuilder* : Construction de ressources lexicales personnelles



# 3. Travaux et développements

- LUCIA

- LUCIAAnalyser* : Projection sur les textes des ressources lexicales personnelles



# 3. Travaux et développements

---

- LUCIA
  - Mise en évidence de faisceaux d'isotopies
    - Filtrages et reclassements personnels sur les réponses de requêtes sur le web
  - Représentations symboliques personnalisées des documents
    - Un repérage rapide des documents intéressants
    - Un repérage des invariants sémiotiques (ex : forme article de Libé)



# 3. Travaux et développements

---

- ProxiDocs (thèse de Thibault Roy)
  - Visualiser un ensemble de documents pour donner à un interprétant une vision globale de la thématique de cet ensemble
  - Visualisation sous forme de « cartes »
    - représentations graphiques topologiques (personnelles)
    - objets supports d'interactions pour l'aide à la lecture intertextuelle

# 3. Travaux et développements

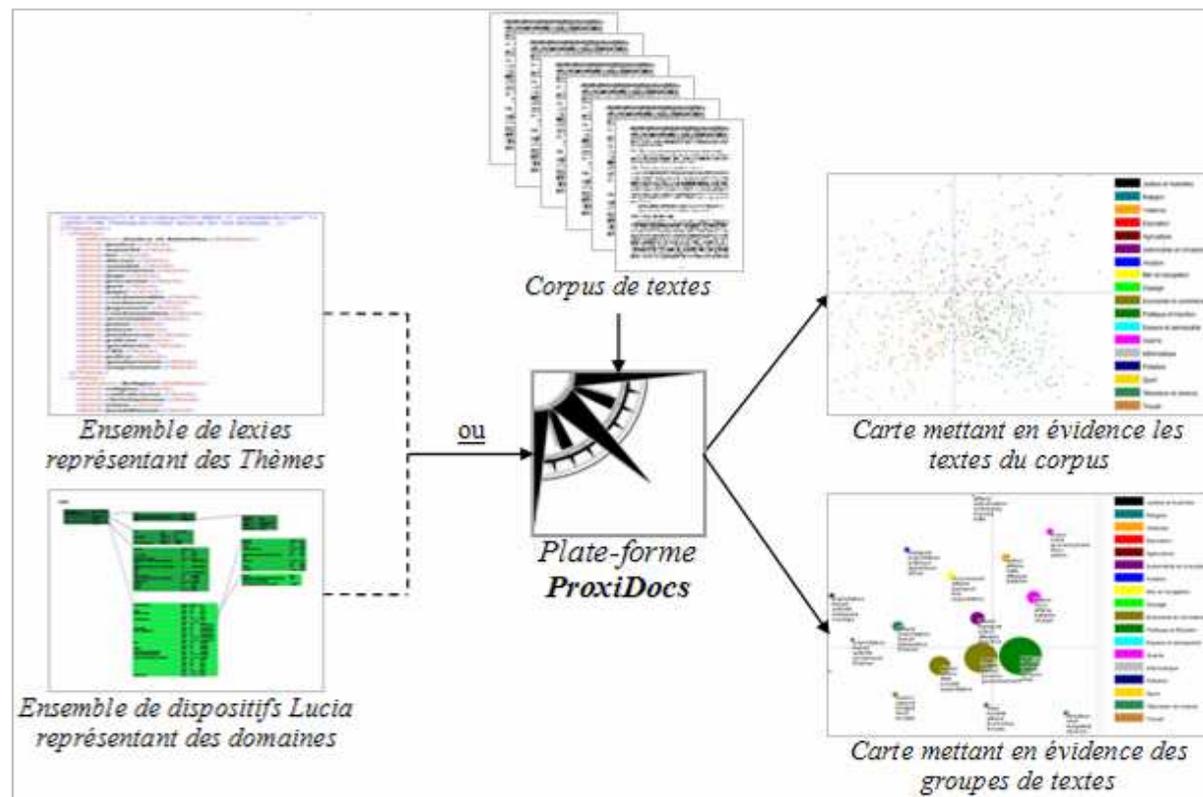
---

- ProxiDocs
  - L'utilisateur peut réaliser diverses manipulations sur les cartes à l'aide de sa souris :
    - Zooms et déplacements
    - Accès aux documents coloriés selon les domaines
    - Accès à des rapports d'analyses de documents et de groupes de documents
    - Accès à des sous-cartes d'un groupe de documents
    - Mise en évidence des documents d'un même domaine
    - Mise en évidence de termes partagés par plusieurs groupes
    - Rotation de l'espace (carte 3D)

# 3. Travaux et développements

- ProxiDocs

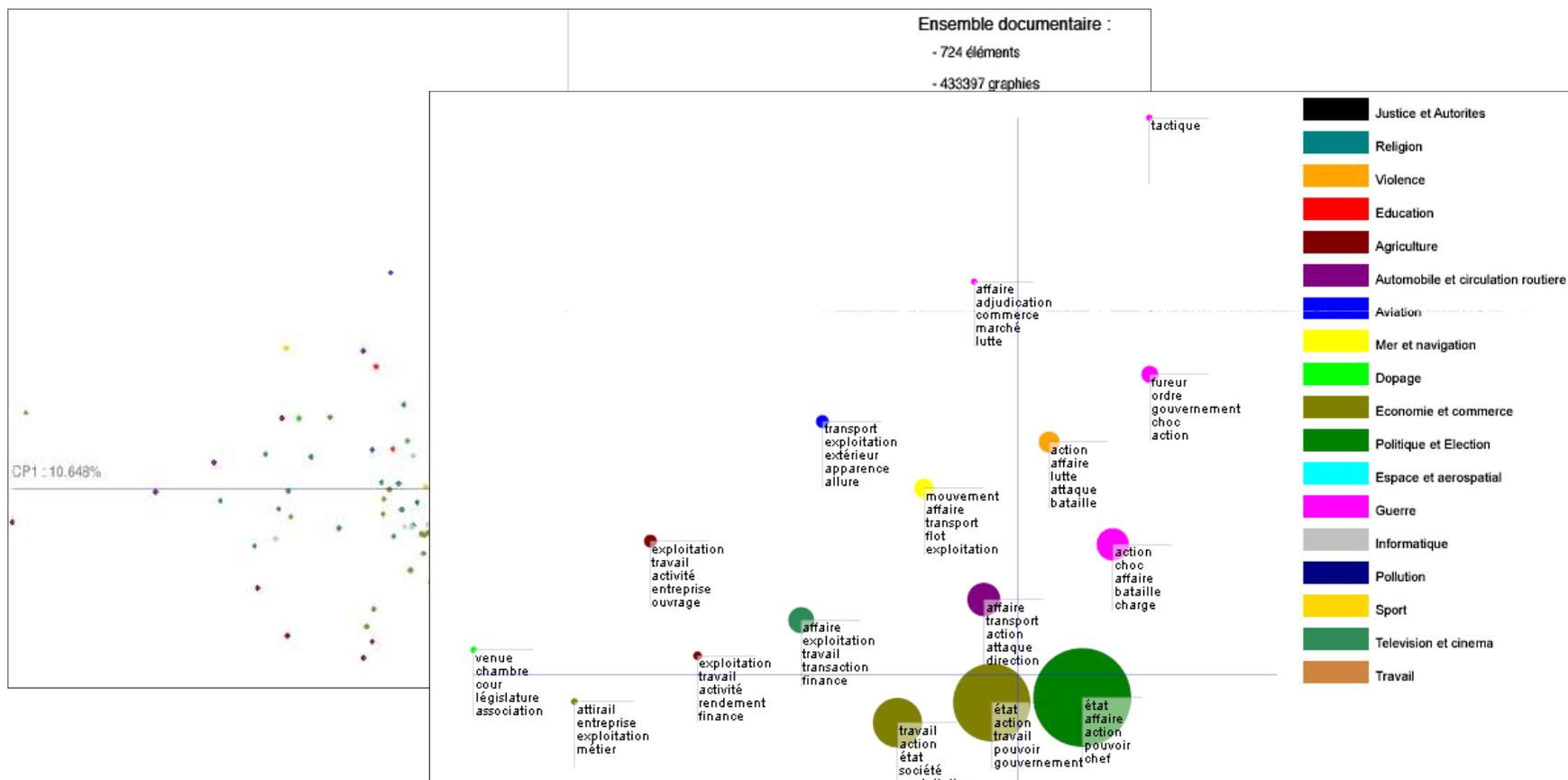
Ressources terminologiques personnelles  
(liste de graphies ou dispositifs différentiels)



Vues globales  
Personnelles  
(cartes de documents, cartes de groupes, ...)

# 3. Travaux et développements

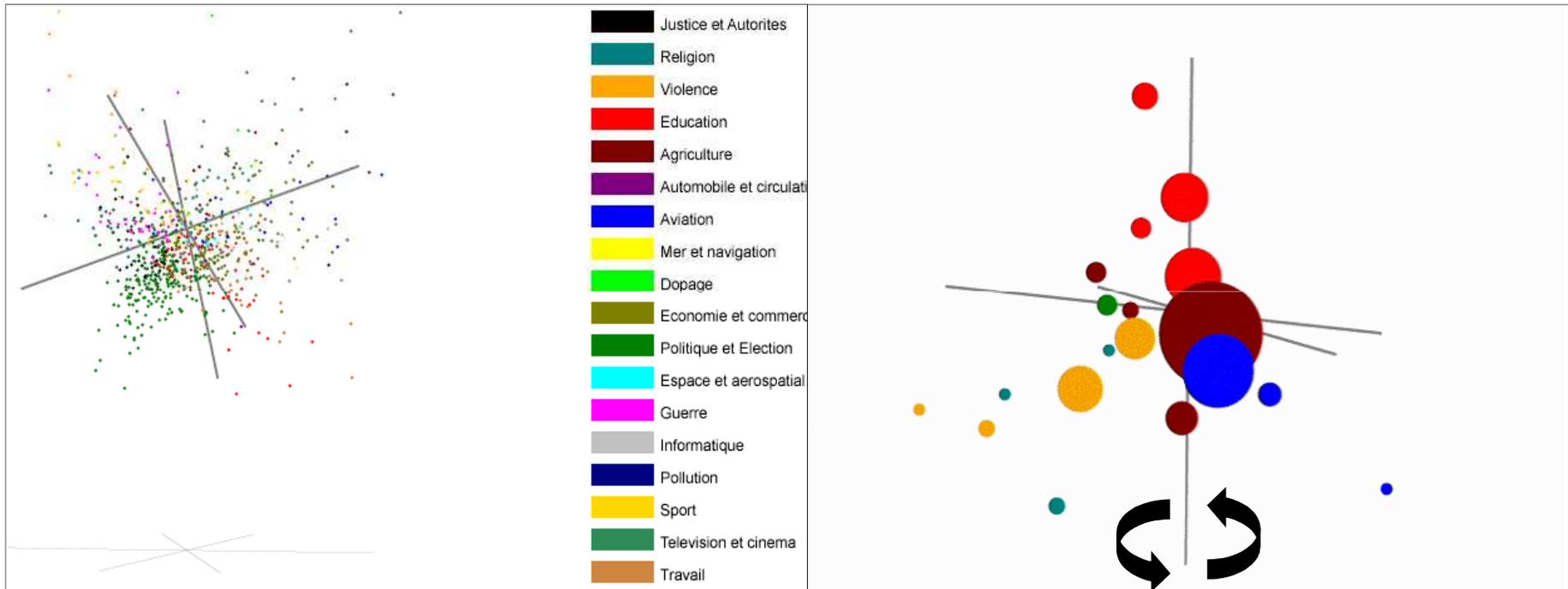
- ProxiDocs



Visualisation de cartes de documents et de carte de groupes de documents

# 3. Travaux et développements

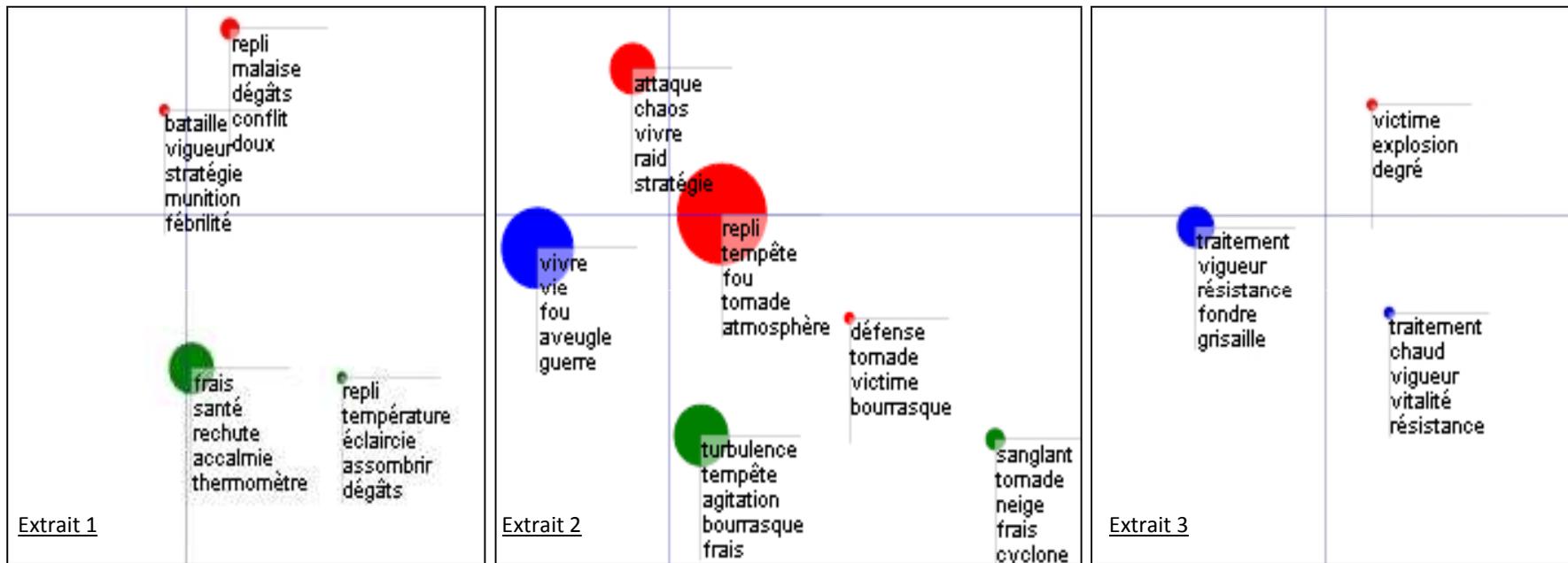
- ProxiDocs



Visualisations en 3D

# 3. Travaux et développements

- ProxiDocs



visualisations dynamiques : fenêtre d'un mois, déplacement d'un jour, actualités du journal Le Monde (1987)

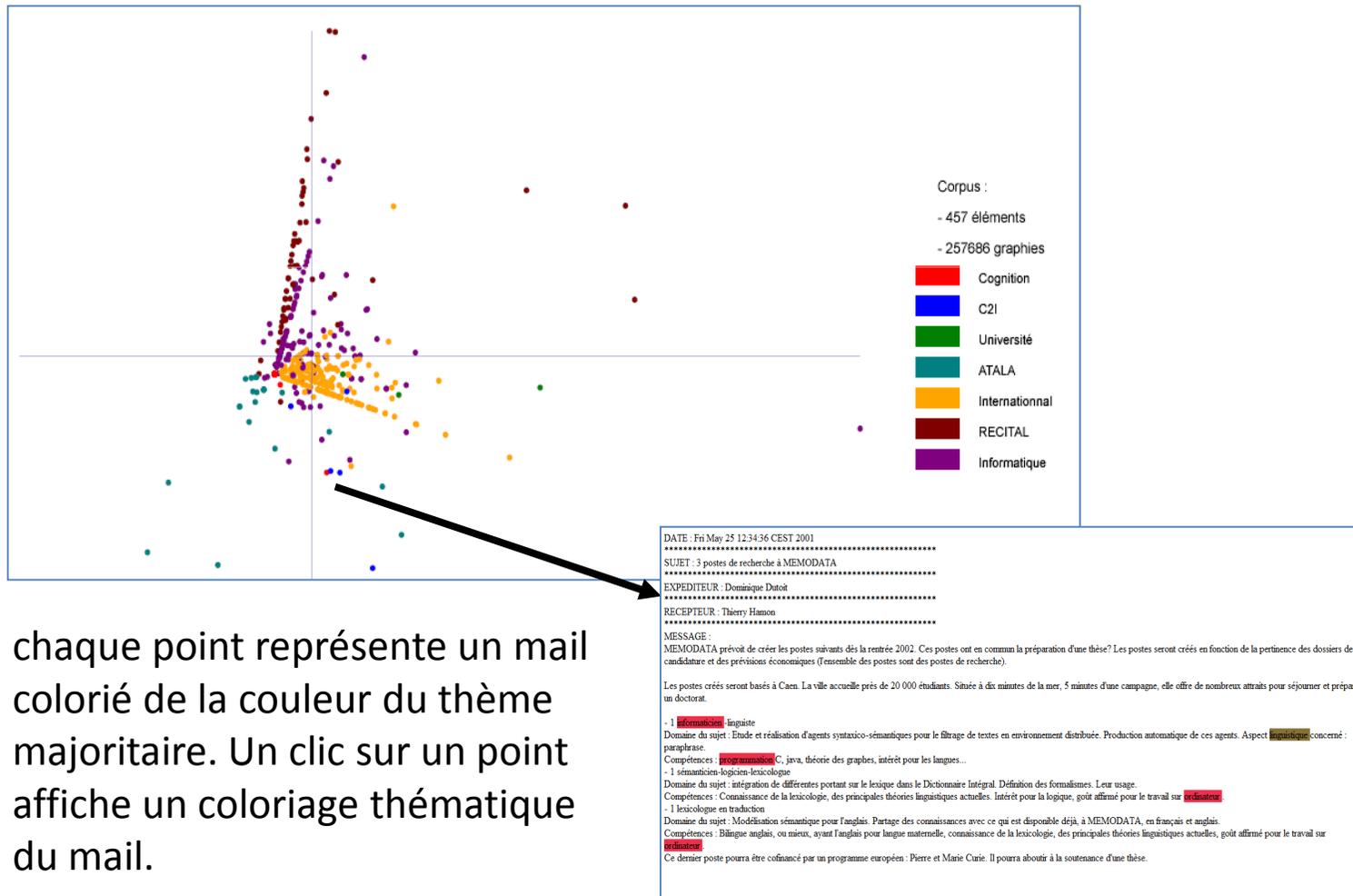
# 3. Travaux et développements

---

- CartoMail
  - Reprendre l'approche de ProxiDocs et ThemeEditor et l'appliquer à des corpus de courriers électroniques
  - Une application qui permette de visualiser de manière interactive et personnelle un compte de messagerie électronique
  - Vision globale et personnalisée de la dimension intertextuelle d'un compte mail (boites, sous-boites et message)
  - Un usage complémentaire aux outils de l'utilisateur et à son plan de classement.
  - Une première maquette (largement buggée !!!) réalisée sous forme de projet d'étudiant de M1 d'informatique.

# 3. Travaux et développements

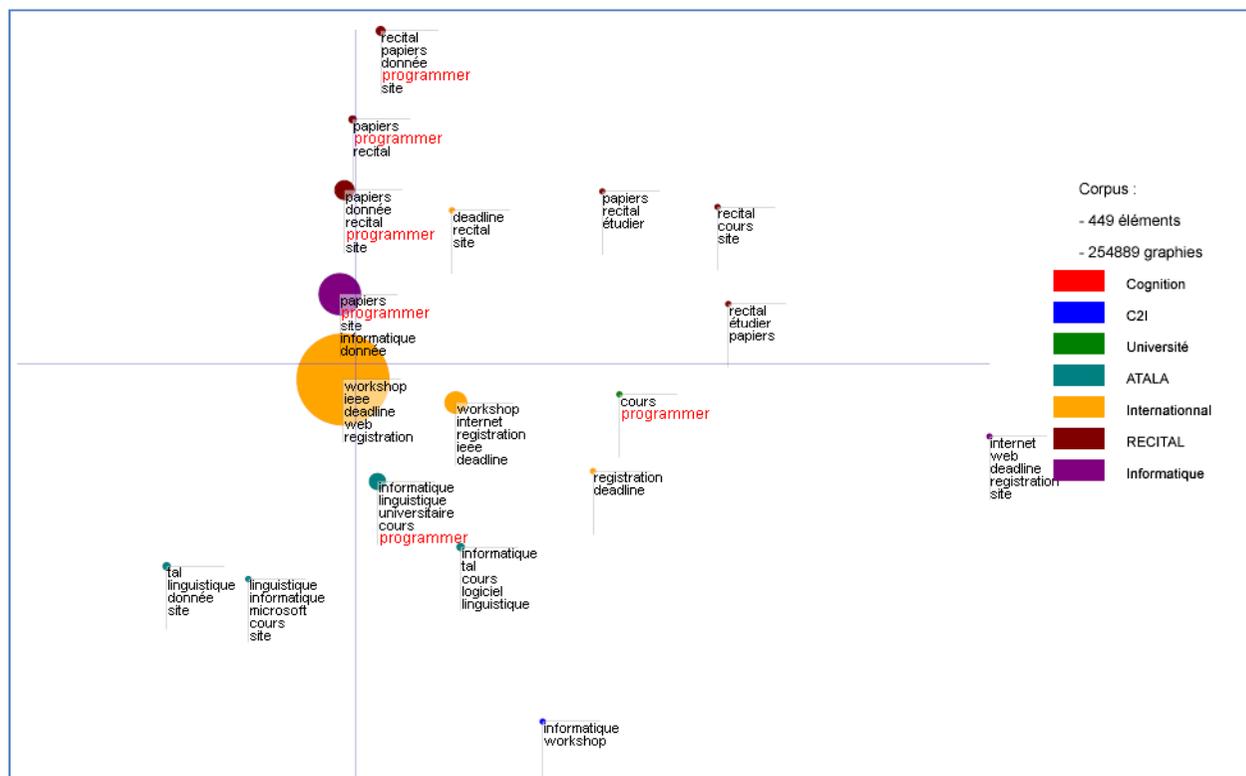
- CartoMail



chaque point représente un mail colorié de la couleur du thème majoritaire. Un clic sur un point affiche un coloriage thématique du mail.

# 3. Travaux et développements

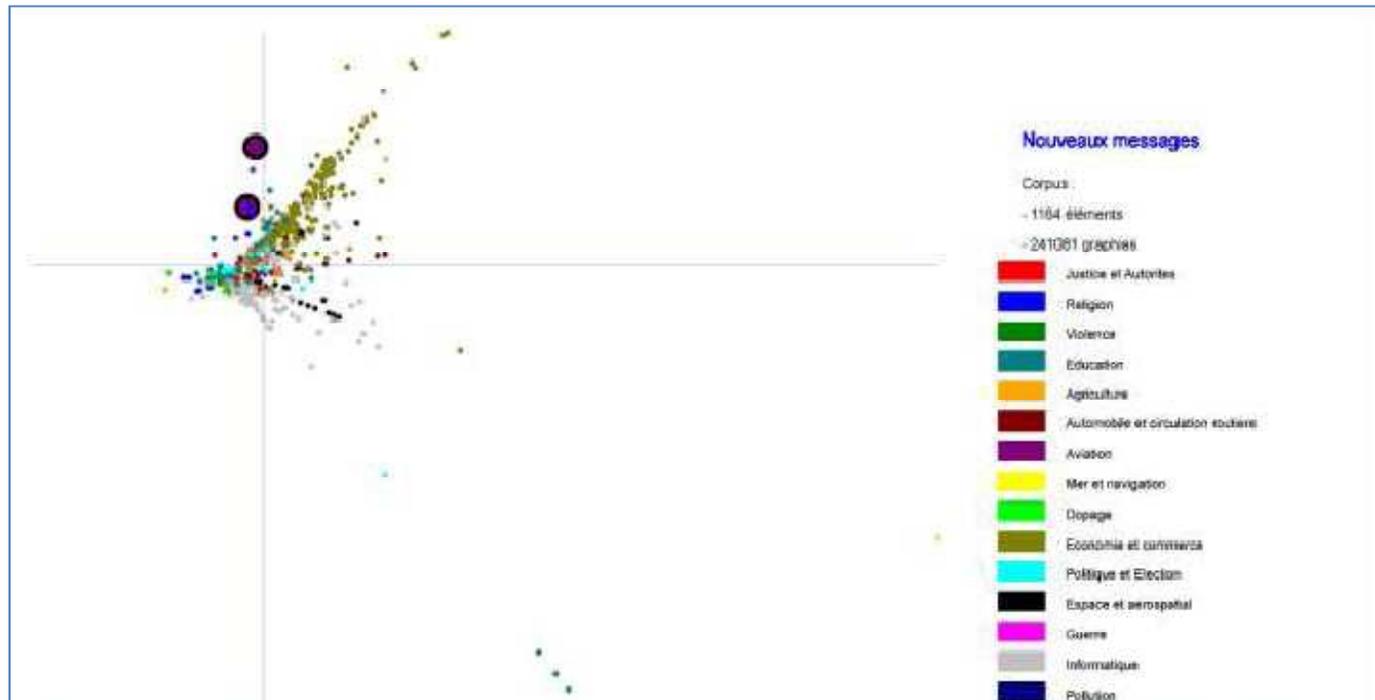
- CartoMail



chaque disque représente un groupe de mails colorié de la couleur du thème majoritaire. Un passage de souris sur un terme affiche en rouge les autres occurrences du terme dans les autres groupes.

# 3. Travaux et développements

- CartoMail



Les points qui apparaissent en plus grosse taille que les autres indiquent la place des nouveaux messages dans l'espace thématique de la messagerie.

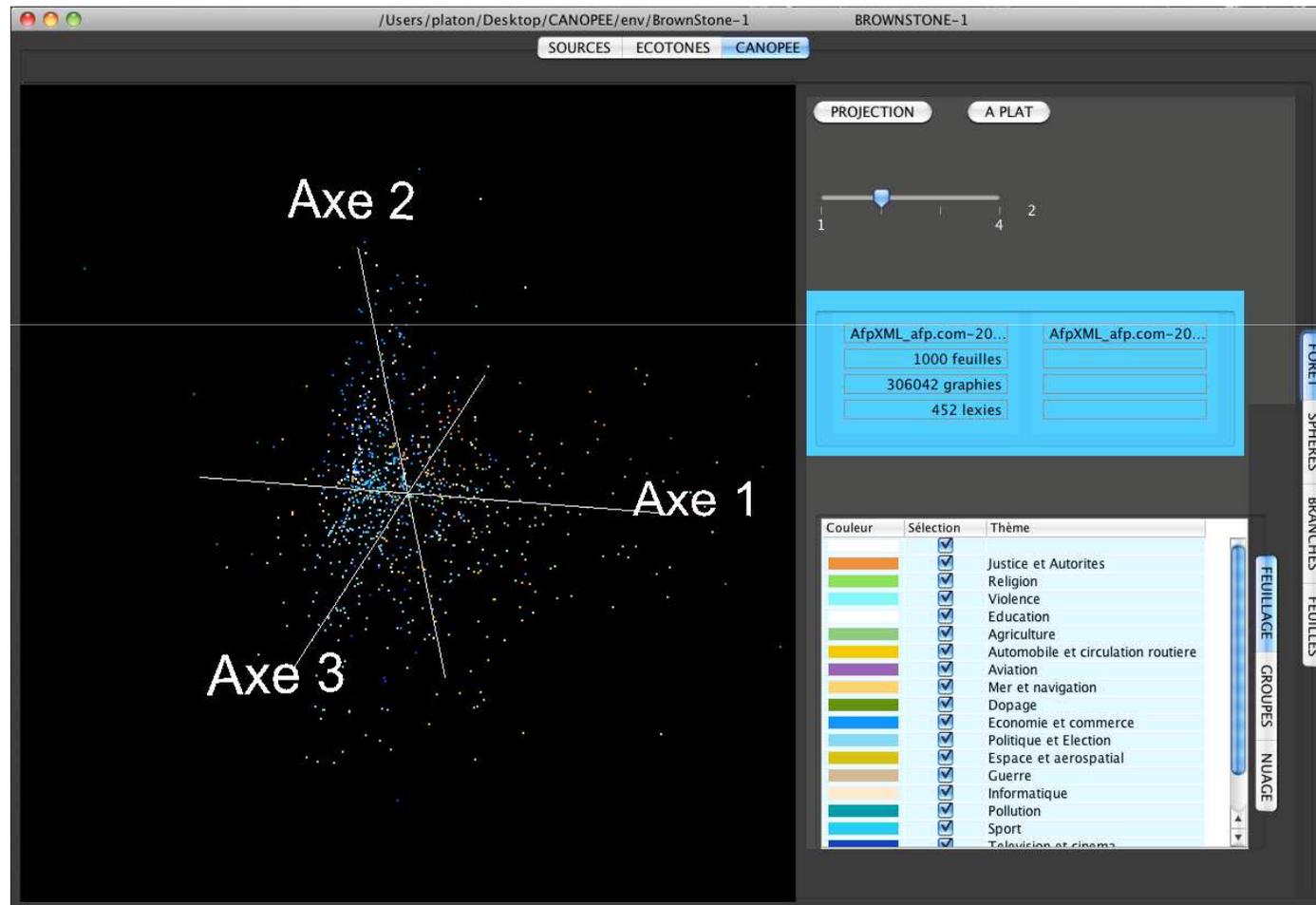
# 3. Travaux et développements

---

- Canopée © ExoMakina
  - Développer une cartographie personnalisée de flux documentaires
  - Aller vers un développement professionnel fiabilisé
  - Assister des veilleurs d'information
    - Cartographie de corpus
    - Cartographie dynamique
    - Cartographie de ressources lexicales
    - Requêtage par l'exemple dans l'espace documentaire
  - Sortie prévue courant 2011

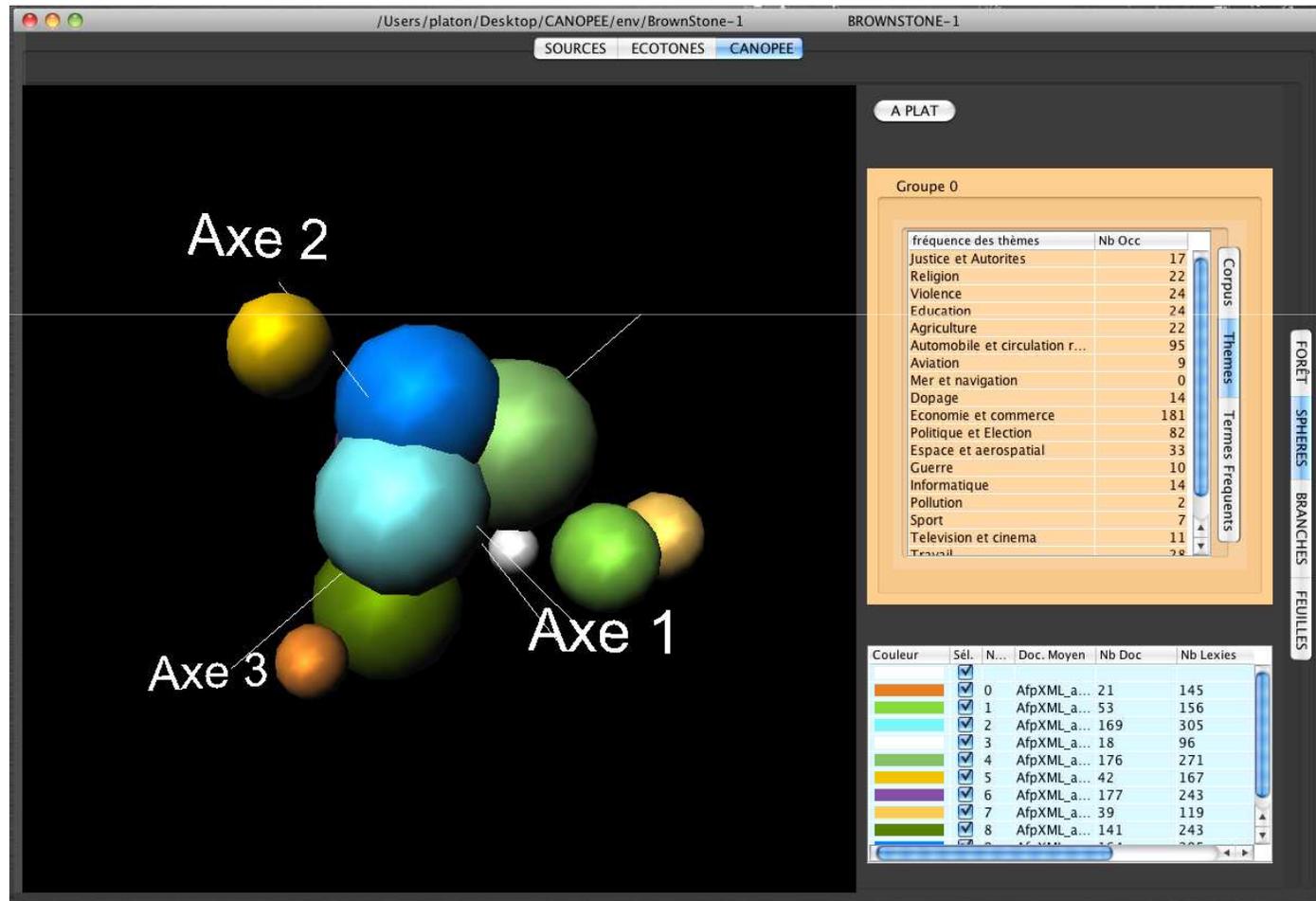
# 3. Travaux et développements

- Canopée © ExoMakina



# 3. Travaux et développements

- Canopée © ExoMakina



# Plan

---

1. Mon parcours
2. Objectifs et problématique
3. Travaux et développements
- 4. Expérimentations sur corpus**
5. Sémantique Interprétative et TAL
6. Vers des ENT énatifs
7. Conclusions

# 4. Expérimentations sur corpus

---

- Plusieurs expérimentations menées
  - Sur différents genres textuels :
    - Analyse des énoncés de dialogue (Anadia)
    - Veille documentaire (LUCIA)
    - Forums de discussion pédagogique (ProxiDocs)
    - Terminologie dans les résumés de thèses (ProxiDocs)
    - Analyses de métaphores conceptuelles boursières (LUCIA+ProxiDocs)
  - Sur les modalités d'usage
    - Tracés oculaires sur les carte ProxiDocs

## 4. Expérimentations sur corpus

---

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) - 2007
  - Utilisation de ProxiDocs
  - 130 résumés de thèses (56 451 graphies)
  - 5 thèmes :
    - Reconnaissance de formes
    - Intelligence artificielle
    - Vision
    - Traitement du signal
    - Traitement d'image

## 4. Expérimentations sur corpus

---

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007
  - Thème « reconnaissance de formes »
- système de perception
- fusion d'informations
- segmentation d'images couleur
- classification
- problèmes inverse
- champs de Markov cachés
- échantillonnage de Gibbs
- tomographie micro-onde
- reconnaissance des formes
- systèmes de classifieurs
- réseaux de Petri flous
- combinaison de classifieurs
- coopération de classifieurs
- sélection de classifieurs
- coalescence floue
- rejet d'ambiguïté
- ensembles flous
- classification non supervisée
- fonction de densité de probabilité
- imagerie multi-modale
- réduction de dimensionalité

## 4. Expérimentations sur corpus

---

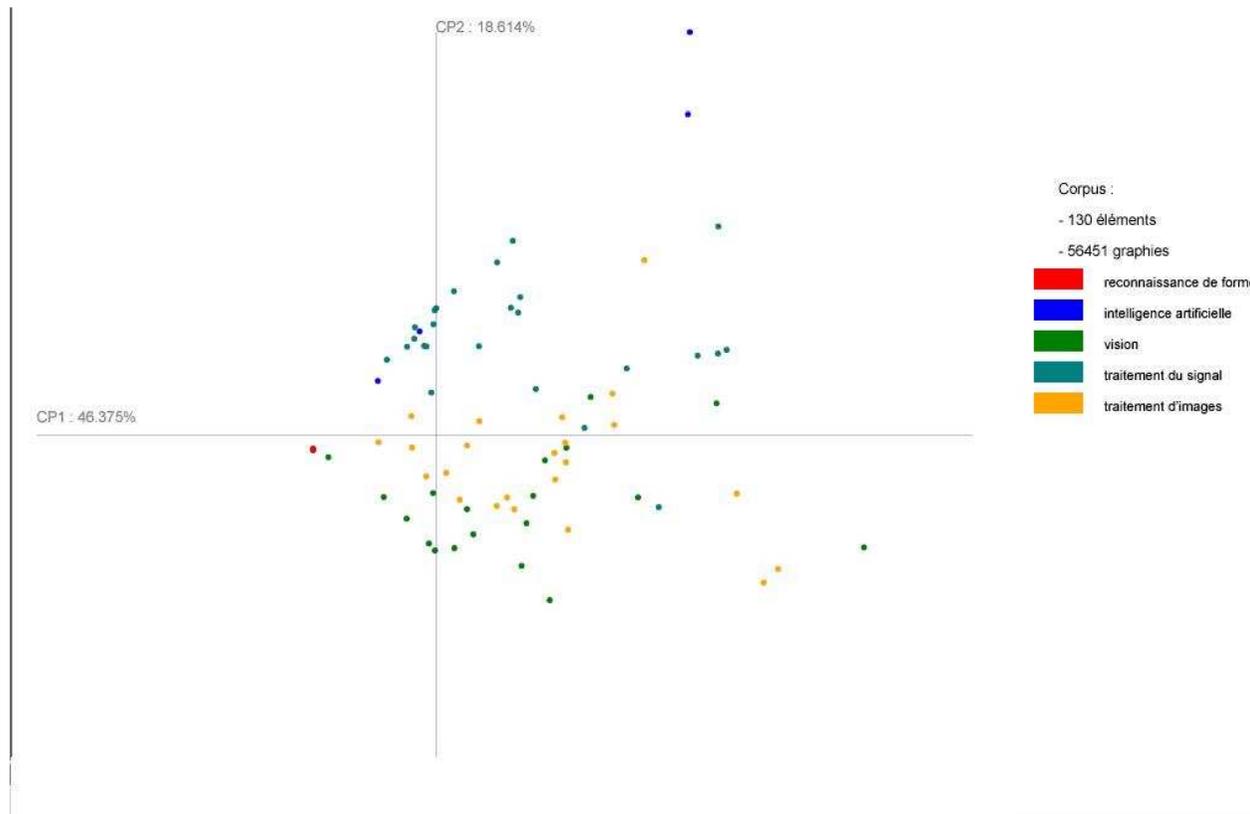
- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007

### – Thème « Intelligence artificielle »

- réseaux de Petri partiellement stochastiques
- diagnostic de pannes dans les systèmes répartis
- algorithmes
- inférence bayésienne
- modèles de Markov cachés
- méthodes de Monte Carlo
- Markov Chain Monte Carlo
- filtrage particulaire
- processus de Dirichlet
- estimation de densité
- fusion de données
- localisation de véhicule
- métaheuristiques
- génie biomédical
- systèmes d'irrigation complexes
- commande optimale
- identification
- modélisation
- programmation
- problèmes inverse
- champs de Markov cachés
- échantillonnage de Gibbs
- tomographie micro-onde
- théorie des croyances
- réseaux de neurones formels
- neurones à impulsions
- détection de cibles
- apprentissage
- détection de signaux radar
- classification
- adéquation algorithmes-architecture
- spécification
- graphes
- implantation
- génération automatique
- théorie de la recherche
- jeux de Reconnaissance
- optimisation
- colonies de fourmis
- gestion des détecteurs et des ressources
- GPS

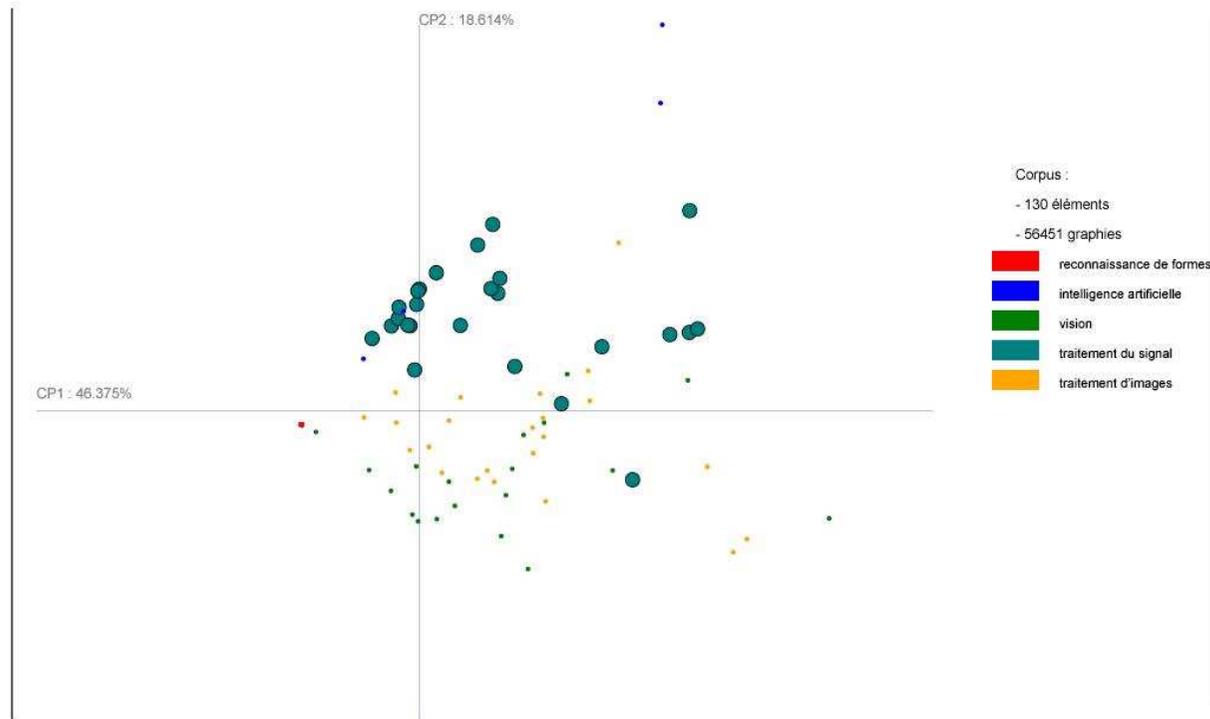
## 4. Expérimentations sur corpus

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007



## 4. Expérimentations sur corpus

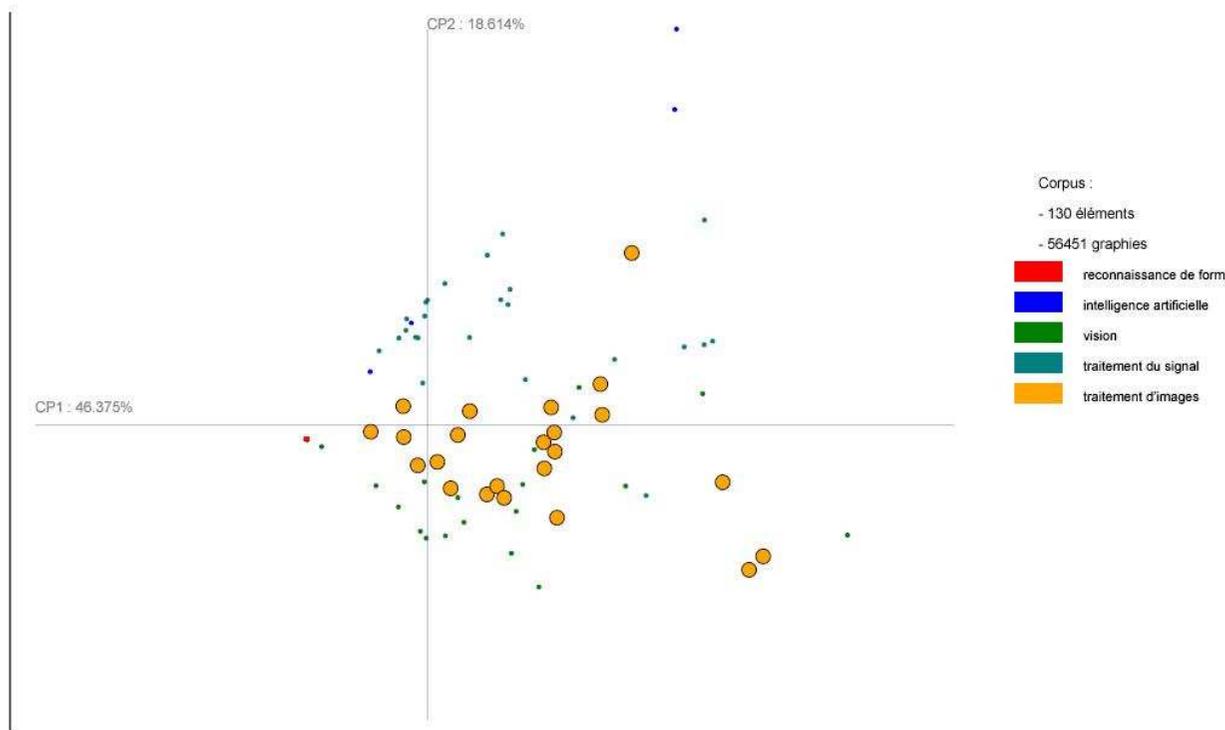
- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007



## 4. Expérimentations sur corpus

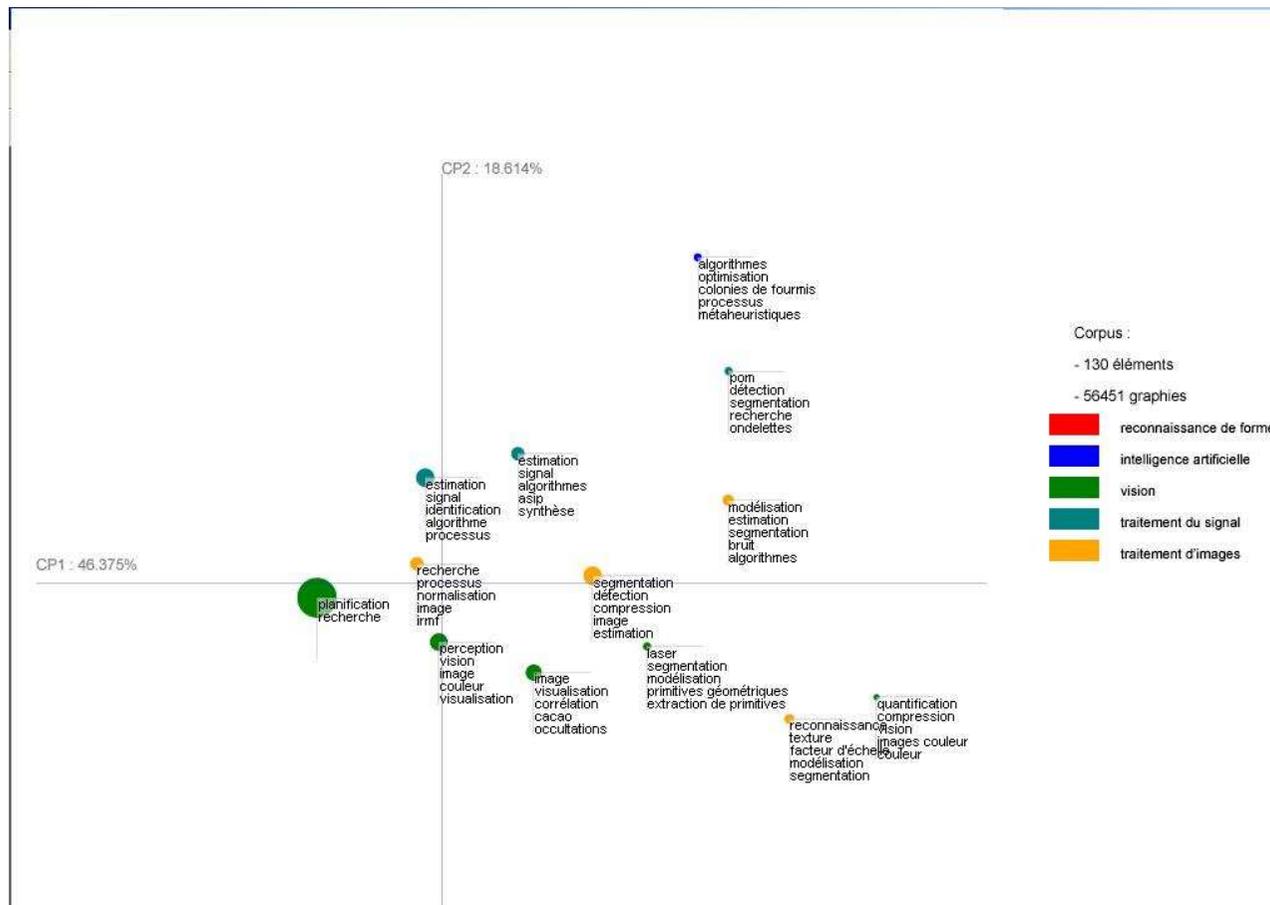
---

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007



# 4. Expérimentations sur corpus

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007



## 4. Expérimentations sur corpus

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007

*Analyse du document original :*  
*/Users/thibaultroy/Documents/FAC/These/enaction/corpus\_resumes\_theses/1.txt*

| Propriétés du document    |          |                  | Propriétés du groupe contenant le document (cardinalité du groupe : 2) |          |  |
|---------------------------|----------|------------------|--|----------|--|
| Domaine                   | Nb. occ. | % occ. du global | Domaine  | Nb. occ. |  |
| reconnaissance de formes  | 0 occ.   | 0.0%             | reconnaissance de formes   | 0 occ.   |  |
| intelligence artificielle | 2 occ.   | 16.6%            | intelligence artificielle  | 12 occ.  |  |
| vision                    | 14 occ.  | 73.6%            | vision   | 19 occ.  |  |
| traitement du signal      | 51 occ.  | 85.0%            | traitement du signal   | 60 occ.  |  |
| traitement d'images       | 19 occ.  | 57.5%            | traitement d'images  | 33 occ.  |  |

| Rang | Lexie               | Nb occ. | Nb doc. supports | % occ. du global | Rang | Lexie               | Nb occ. | Nb de supports |
|------|---------------------|---------|------------------|------------------|------|---------------------|---------|----------------|
| 1.   | <u>pom</u>          | 7 occ.  | 1 doc.           | 100.0%           | 1.   | <u>pom</u>          | 7 occ.  | 1 doc          |
| 2.   | <u>segmentation</u> | 5 occ.  | 1 doc.           | 100.0%           | 2.   | <u>détection</u>    | 6 occ.  | 2 doc          |
| 3.   | <u>ondelettes</u>   | 4 occ.  | 1 doc.           | 100.0%           | 3.   | <u>segmentation</u> | 5 occ.  | 1 doc          |
| 4.   | <u>parole</u>       | 4 occ.  | 1 doc.           | 100.0%           | 4.   | <u>recherche</u>    | 5 occ.  | 1 doc          |
|      |                     |         |                  |                  | 5.   | <u>ondelettes</u>   | 4 occ.  | 51 doc         |

## 4. Expérimentations sur corpus

---

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007

Document original colorié : ABDALLAH imad : REPRESENTATIONS TEMPS-FREQUENCE ADAPTATIVES DE SIGNAUX ACOUSTIQUES BASEES SUR DES CRITERES ENTROPIQUES

Thèse ou HDR soutenue le 11/12/1998 à Institut d'Informatique Claude

Contact : silvio@lium.univ-lemans.fr Jury M. BAUDRY Directeur, Professeur, LIUM, Le Mans C. DONCARLI Rapporteur, Professeur, IRCYN, Nantes D. OUAHABI Rapporteur, Professeur, EIT, Tours C. D'ALESSANDRO Examineur, CR Habilité, LIMSI CNRS, Orsay M.F. LUCAS Examineur, MdC, IR

Direction Marc Baudry Silvio Montrésor

Laboratoire Laboratoire d'Informatique de l'Université du Maine

Résumé L'objectif de ce travail est de concrétiser l'idée déjà ancienne que la description d'un signal non stationnaire par une base de décomposition est, en terme de lisibilité, d'autant meilleure que celle ci peut s'adapter aux évolutions du signal. Nous étudions dans un contexte unifié, qui repose sur la notion des bancs de filtres, les bases orthogonales dyadiques issues des Paquets d'Ondlettes de Malvar (POM), et des Paquets d'Ondelentes (PO). Les applications étudiées ont pour cadre les traitements des signaux acoustiques et plus particulièrement la segmentation en vue de la compression, la réduction de bruit et la reconnaissance de la parole. Le premier apport de cette thèse est de proposer des extensions des décompositions dyadiques en POM et en PO. Ces extensions, que nous appelons s-dyadiques, permettent d'éliminer les ruptures artificielles sur les contenus temporel et fréquentiel du signal qui sont inhérentes aux structures dyadiques. Elles offrent une plus grande liberté de choix des pavages temps-fréquence associés aux POM et aux PO. Nous proposons deux algorithmes basés sur des critères entropiques et énergétiques qui permettent de choisir les décompositions s-dyadiques adaptées. Le deuxième apport concerne une méthode originale de segmentation basée sur des critères entropiques et qui s'apparente à une analyse multirésolution de type POM à 2 niveaux. Le point commun de ces différentes approches est l'utilisation de critères issus de l'entropie de Shannon pour la sélection des bases adaptées. Pour la validation de ces approches, nous avons d'abord développé un algorithme de segmentation/compression des signaux de parole fondé sur les POM s-dyadiques. Cet algorithme offre des taux de compression plus élevés que ceux issus de la DCT, des transformées à recouvrement orthogonales uniformes et des POM dyadiques. Deuxièmement, nous avons effectué une comparaison montrant que les méthodes de réduction de bruit qui utilisent les représentations s-dyadiques adaptées en POM et en PO sont plus performantes que celles fondées sur la transformée en ondélentes discrète, les POM et les PO dyadiques. Enfin, nous avons montré que pour la segmentation des signaux de parole, les algorithmes issus des critères entropiques permettent d'effectuer une

## 4. Expérimentations sur corpus

---

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007

### *Détails du groupe n°9*

Nombre de documents du groupe : 12 sur 130 (soit 9.2% du global)

Nombre de graphies dans les documents du groupe : 4494 sur 56451 (soit 7.9% du global)

Nombre de lexies des ressources présentes dans les documents du groupe : 83

Le document "moyen" du groupe :

*[../.../corpus/resumes/theses/rapports/rapport\\_12.txt.html](http://.../corpus/resumes/theses/rapports/rapport_12.txt.html)*

Accès direct aux éléments du rapports :

- [La répartition des catégories dans les documents du groupe](#)
- [Les lexies des catégories présentes dans les documents du groupe](#)
- [Le détail des documents du groupe](#)

## 4. Expérimentations sur corpus

- Terminologie dans un corpus de résumés de thèses (collaboration avec Nathalie Beaudouin) – 2007

Répartition des domaines dans les documents du groupe :

| Domaine                   | Nb. occ. |
|---------------------------|----------|
| reconnaissance de formes  | 4 occ.   |
| intelligence artificielle | 26 occ.  |
| vision                    | 115 occ. |
| traitement du signal      | 135 occ. |
| traitement d'images       | 169 occ. |

[\(haut du rapport\)](#)

Répartition des domaines dans le corpus :

| Domaine                   | Nb. occ. |
|---------------------------|----------|
| reconnaissance de formes  | 24 occ.  |
| intelligence artificielle | 199 occ. |
| vision                    | 602 occ. |
| traitement du signal      | 689 occ. |
| traitement d'images       | 712 occ. |

## 4. Expérimentations sur corpus

---

- Etudes des métaphores conceptuelles dans l'actualité boursière (collaboration avec Stéphane Ferrari)
  - Quelle est la dynamique sémique mise en évidence dans l'interprétation des métaphores ?
  - Comment caractériser l'apport interprétatif (nouveau ou analogie) de certaines métaphores conceptuelle en terme de « transferts » de sèmes ?
  - Corpus :
    - Le Monde sur CD-ROM – 1987/1989 plus de 500 articles (500 000 mots)

## 4. Expérimentations sur corpus

- Etudes des métaphores conceptuelles dans l'actualité boursière
  - Interprétation par analogie

Extrait de l'article n°126

Le **Dow Jones** par exemple, le **thermomètre** de la **Bourse** de New York, qui avait chuté de 508 points (...).

| Objets du domaine           | Rôle           |
|-----------------------------|----------------|
| ciel, pression, température | intervient     |
| degrès, bar                 | étude, analyse |

| Instruments de mesure | Rôle           |
|-----------------------|----------------|
| anémomètre            | vent           |
| pluviomètre           | précipitations |
| thermomètre           | température    |
| mercure, baromètre    | pression       |

Extrait du dispositif en rapport avec la météorologie  
(les tables ont été tronquées).

| Objets du domain                       | Rôle           |
|--|----------------|
| marché, cours ...                      | intervient     |
| graphiques, ratio, courbes, indices... | étude, analyse |

| Indicateurs boursiers | Contexte géographique |
|-----------------------|-----------------------|
| CAC, CAC40            | France                |
| Dow Jones, Nasdaq     | U.S.A.                |
| Nikkei                | Japon                 |
| Dax                   | Allemagne             |
| Footse                | Royaume Uni           |

Extrait du dispositif en rapport avec la bourse  
(les tables ont été tronquées).

## 4. Expérimentations sur corpus

- Etudes des métaphores conceptuelles dans l'actualité boursière
  - Interprétation par nouveauté

Extrait de l'article n°126

Ce **krach** était dû (...) à la chute vertigineuse et incontrôlée du **dollar**, signe que la **tempête** affecte dorénavant les **marchés financiers**.

| Phénomènes dynamiques                      | Direction | Evaluation  |
|--|-----------|-------------|
| dépréciation, dévaluation, krach, dévaluer | descend   | mauvais     |
| baisse des cours, inflation                | monte     | pas connoté |
| hausse des cours, déflation                | descend   | pas connoté |

Extrait du dispositif en rapport avec la bourse  
(les tables ont été tronquées).

| Phénomènes dynamiques | Direction | Evaluation | Axe         |
|-----------------------|-----------|------------|-------------|
| gel, geler            | descend   | mauvais    | température |
| intempéries           | monte     | mauvais    | temps       |
| accalmie              | descend   | bon        | temps       |

| Intempéries     | Force        |
|-----------------|--------------|
| rafale, tempête | violent      |
| cyclon, typhon  | très violent |

Extrait du dispositif en rapport avec la météorologie  
(les tables ont été tronquées).

Sème actualisable dans les isotopies issues de la co-textualité de *krach* et *tempête* <sup>57</sup>

## 4. Expérimentations sur corpus

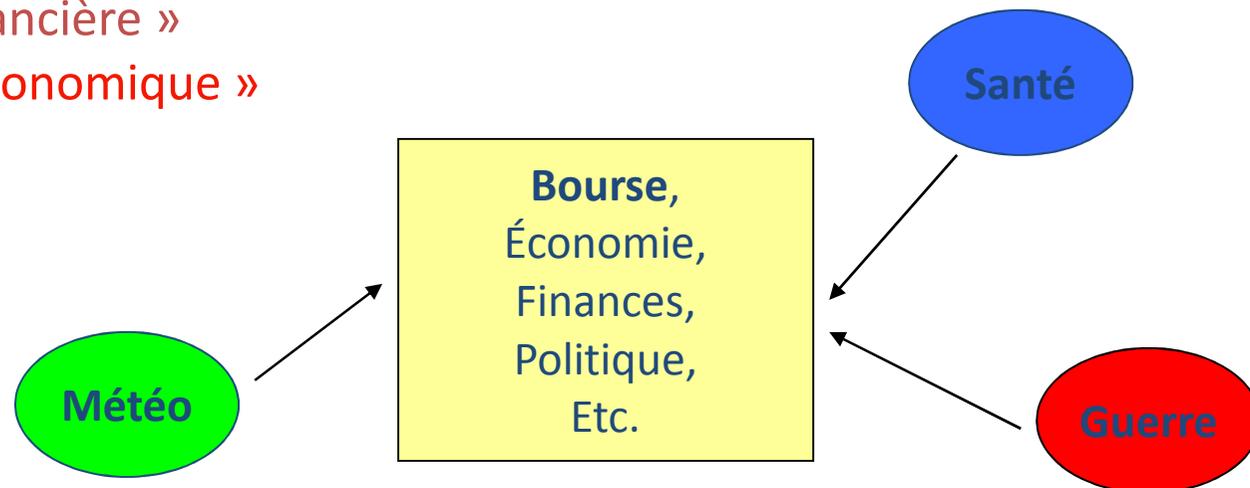
---

- Etudes des métaphores conceptuelles dans l'actualité boursière
  - Ressources : 3 dispositifs différentiels (64+112+111 lexies)
  - Observations : 3 métaphores conceptuelles

La « météorologie boursière »

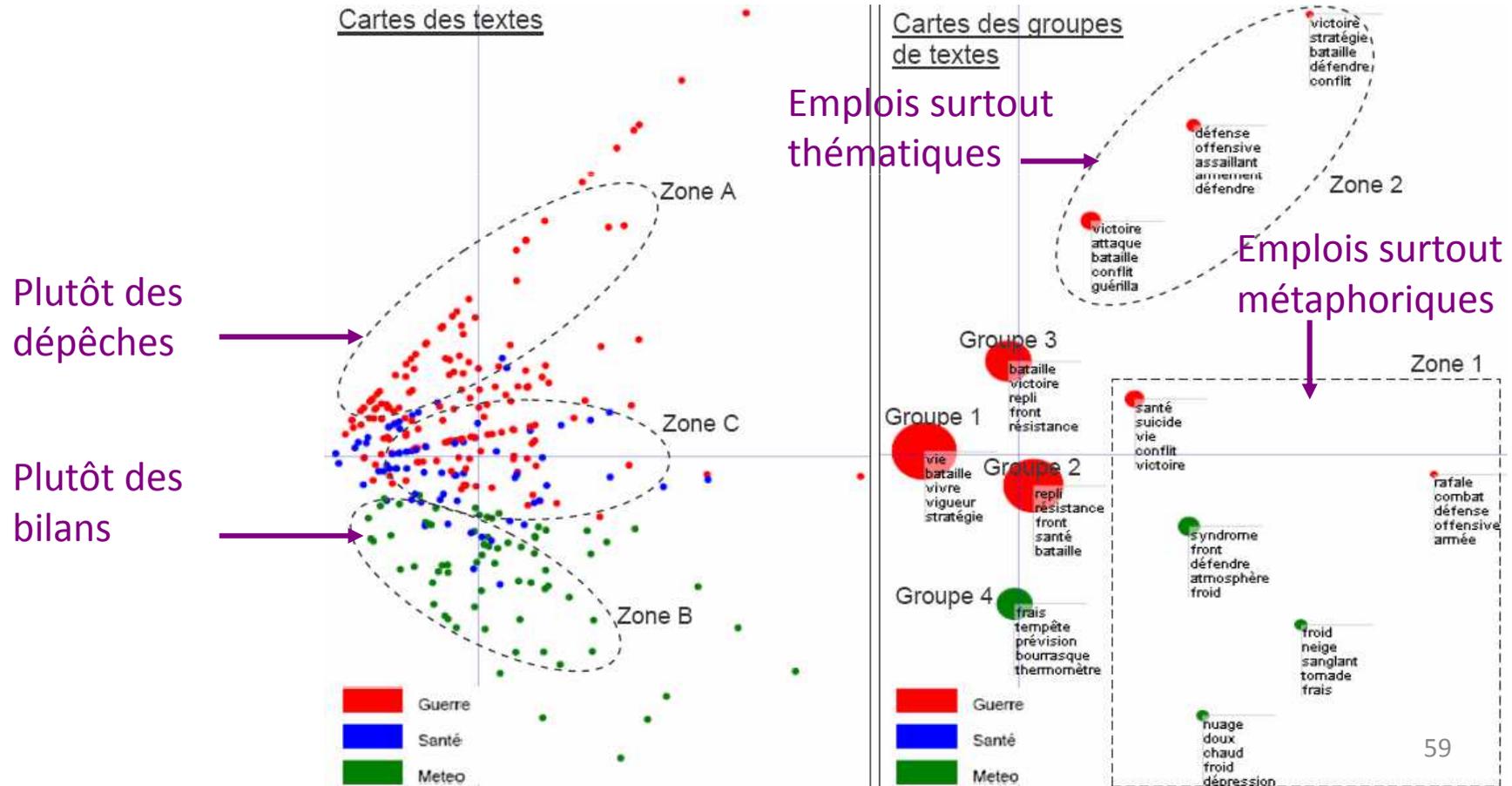
La « santé financière »

La « guerre économique »



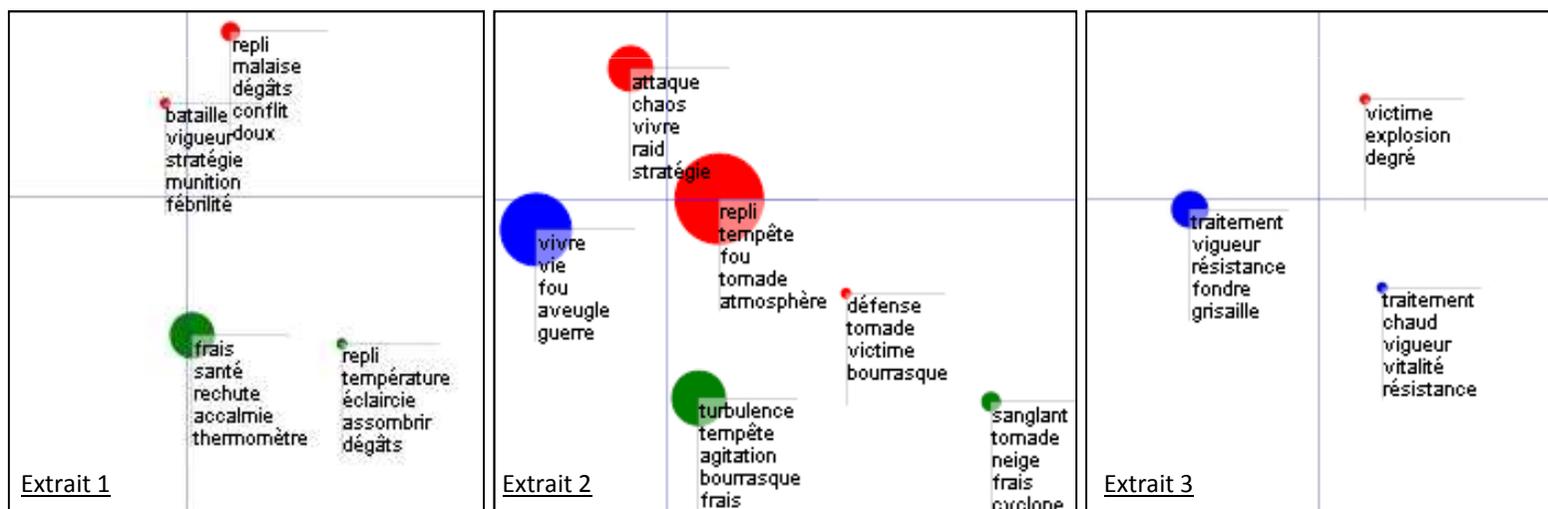
# 4. Expérimentations sur corpus

- Etudes des métaphores conceptuelles dans l'actualité boursière : analyse des cartes



## 4. Expérimentations sur corpus

- Etudes des métaphores conceptuelles dans l'actualité boursière : évolutions diachroniques (fenêtre d'un mois, déplacement d'un jour)



# Plan

---

1. Mon parcours
2. Objectifs et problématique
3. Travaux et développements
4. Expérimentations sur corpus
- 5. Sémantique Interprétative et TAL**
6. Vers des ENT éactifs
7. Conclusions

# 5. Sémantique Interprétative et TAL

---

- Paradigme dominant en TAL : la formalisation du sens
  - Tradition logico-grammaticale décrite par F. Rastier
  - Depuis la fin des années 50 le TAL a principalement adopté une vision très peu interactive du langage. La langue est vue comme un objet théorisable, et donc qu'on peut « mettre en machine ».
  - C'est l'influence du cognitivisme qui voit la pensée et la communication comme des processus de traitement de l'information
  - Approches « computo-représentationnelles » des langues et du langage (Chomsky, Fodor ...)
  - La signification est un consensus représentable

Le « tout théorique », « tout formalisable », « tout ontologique », « tout automatique »

# 5. Sémantique Interprétative et TAL

---

- L'exemple du Web sémantique : surcharger les documents avec « la » représentation de leur sens utilisable par des machines
  - Le sens serait décrit de manière consensuelle ! (domination culturelle)
  - Le sens serait déterminé par le point de vue de l'auteur du document !
  - Le sens est indépendant de tout contexte d'interprétation !
  - Le sens est construit à partir de bases de connaissances ontologiques (type RDF) de manière compositionnelle !
  - Le sens est une donnée (qui se passe du texte) !

cf. Le Web des données : Linked Data

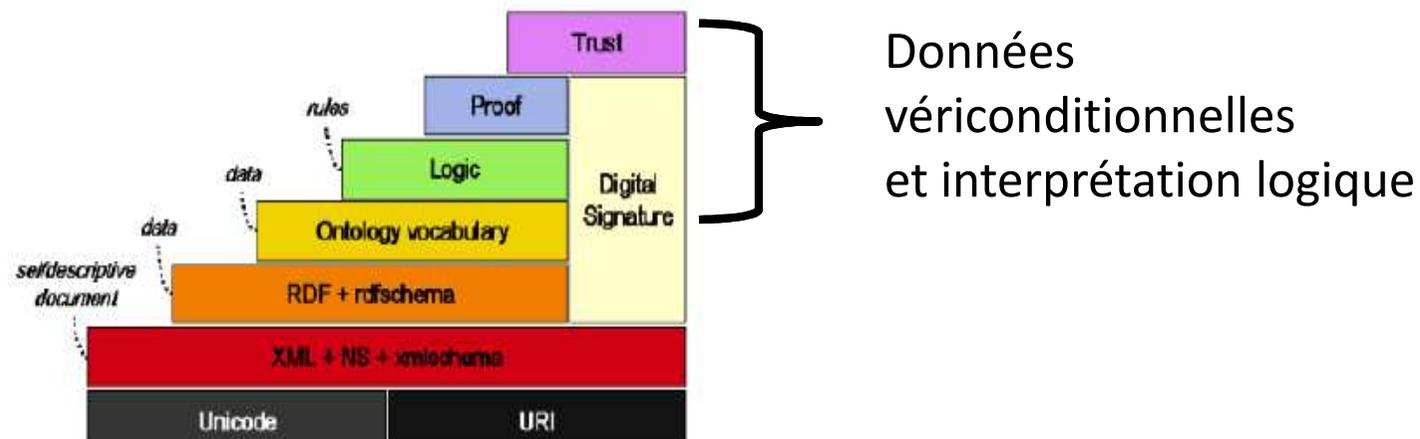
Voir la vidéo de Tim Berners-Lee:  
The next Web of open, linked data (13 mars 2009)

[http://www.youtube.com/watch?v=OM6XIICm\\_qo](http://www.youtube.com/watch?v=OM6XIICm_qo)



# 5. Sémantique Interprétative et TAL

- Après le web des documents, le web des données
- Linked Data : les données, c'est le sens
  - Hors des textes
  - Hors du contexte
- Se passer de la dimension textuelle du sens
- Le sens ça doit servir à quelque chose (i.e. des applications sur le web)



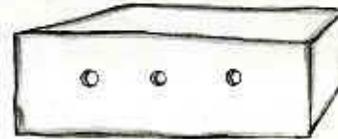
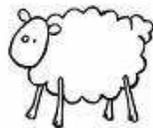
The W3C Semantic Web Layer Cake

# 5. Sémantique Interprétative et TAL

---

- L'exemple du Web sémantique : problèmes !
  - Ontologies uniquement cumulatives
  - Pas de prise en compte de la détermination du local par le global
  - Pas de prise en compte des variabilités d'interprétation
  - Pas de prise en compte des effets contextuels
- Changer de cadre théorique et épistémologique
  - Rejeter une vision ontologique de la signification

« Dessines moi un *mouton* »



- Rejeter une vision « utilitariste » du sens (on ne se sert pas de la langue, c'est notre milieu)
- Préférer l'activité de contextualisation à une « représentation » du contexte
- Subjectiver plutôt qu'objectiver ...

# 5. Sémantique Interprétative et TAL

---

- Ressources subjectives (approche anthropocentrée) :
  - Exemple de la *folksonomie* dans les applications Web 2.0
    - Un tag est une étiquette ou un mot-clé : balise sémantique qui permet de qualifier un contenu ( $\approx$  un sème)
    - Ceux qui taggent, sont ceux qui interprètent ...
    - Un tag n'est pas forcément un concept. Exemples :
      - navigation (concept)
      - à lire, bof , top (rien a voir avec un concept)
    - Etiquetage social qui sert en retour pour alimenter des services de recherche d'information
      - Exemple : Related Tag Browser : outil pour rechercher et naviguer dans les photos publiées sur Flickr à partir de leurs tags
  - De l'anthropocentré au socio-centré

# 5. Sémantique Interprétative et TAL

---

- La sémantique interprétative comme alternative aux approches logico-grammaticales
- Apports de la sémantique interprétative :
  - « appareillage théorique » :
    - La modélisation sémique
    - La dynamique des sèmes (isotopies génériques, spécifiques, actualisations, virtualisations)
    - Les opérations interprétatives (afférences, assimilations, dissimilations ...)
  - « background » épistémologique :
    - Approche non ontologique de la signification
    - Subjectivité et caractère situé de l'interprétation
    - Importance des rapports fond-forme
    - Détermination du global sur le local
    - Pas de réductions simplificatrices du sens (le sens est une différence, dans des parcours interprétatifs et en synchro-diachronie)
    - Une approche contextuelle qui ne suppose pas une représentation du contexte
    - Sciences de la culture

# 5. Sémantique Interprétative et TAL

---

- Le sens d'un document dépend de son interprétant (anagnose, cf. T. Thlivitis) et de son environnement textuel (même corpus, lectures différentes qui s'enrichissent mutuellement, sujets favoris de l'utilisateur ...)
- C'est la collection documentaire qui influe sur le document plus que le contraire
  - Le rapport du global sur le local
  - Les échelles de textualité
  - Les niveaux de contexte d'interprétation
- Pour une caractérisation de l'espace interprétatif

# 5. Sémantique Interprétative et TAL

---

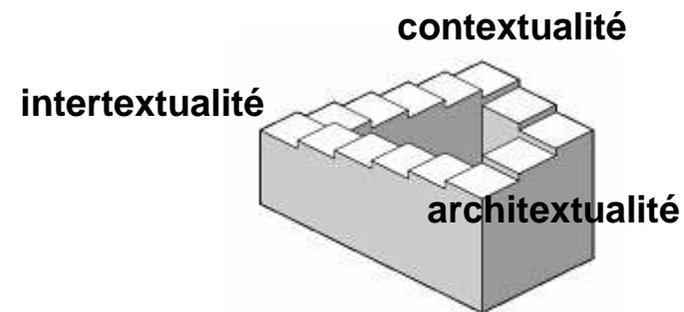
- Les échelles de textualité :
  - le principe de **contextualité**
    - Deux **signes** ou deux passages d'un même texte mis côte à côte sélectionnent réciproquement des éléments de signification (sèmes). Cette sélection réciproque s'analyse notamment en termes de récurrences de sèmes appelées isotopies
  - le principe d'**intertextualité**
    - Deux **passages de textes** différents sélectionnent réciproquement dès qu'ils sont mis côte à côte, des éléments de signification (sèmes). Le principe d'intertextualité généralise la notion d'isotopie au-delà de la linéarité syntagmatique du texte
  - le principe d'**architextualité**
    - Tout **texte** placé dans un corpus en reçoit des déterminations sémantiques, et modifie potentiellement le sens de chacun des textes qui le composent.

# 5. Sémantique Interprétative et TAL

---

- Les échelles de textualité :
  - Le principe de contextualité est à la base des deux autres principes mais il n'est pas autant une étape préalable
  - L'appréhension des rapports « global – local » dans l'interprétation procède à tous les niveaux de contextualité, d'intertextualité et d'architextualité d'alternances entre les formes et les fonds.

Echelles circulaires :



# 5. Sémantique Interprétative et TAL

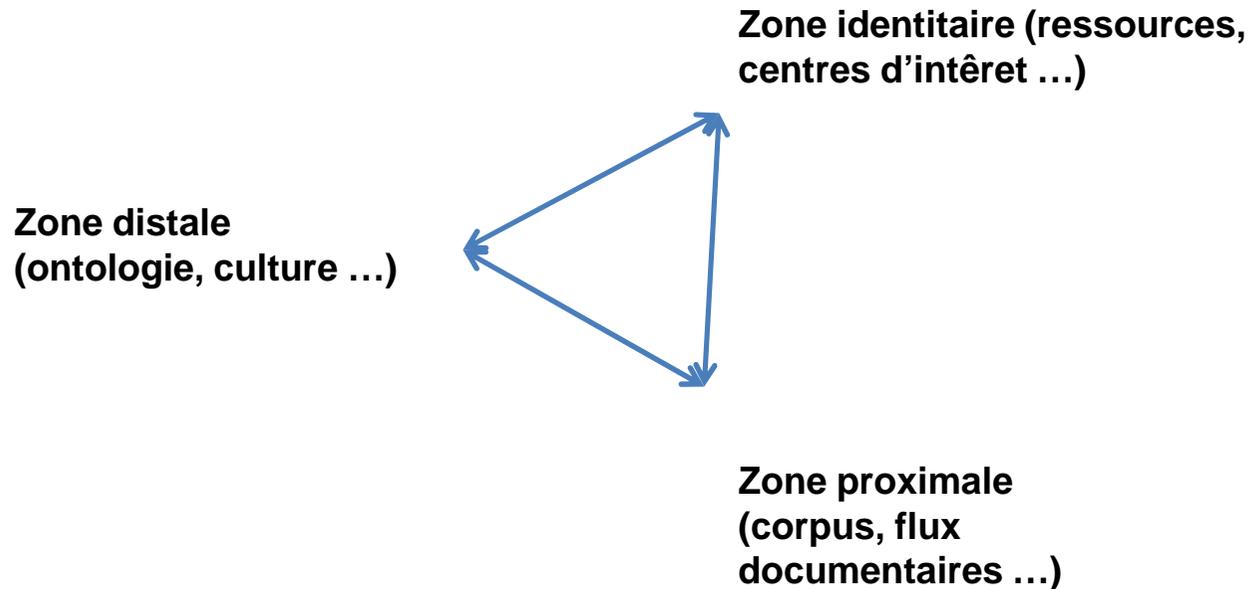
---

- Les niveaux de contexte d'interprétation :
  - La zone identitaire :
    - C'est là que s'exprime **l'individualité**, la subjectivité, les points de vue particuliers. Textuellement, c'est la zone de la première personne verbale (je, nous) dans l'ici et le maintenant.
  - La zone proximale :
    - C'est l'environnement linguistique immédiatement proche du sujet. C'est là que se trouve la communication avec **l'autre**. Textuellement, c'est la zone marquée par la deuxième personne verbale (tu, vous), par le passé proche et futur immédiatement probable.
  - La zone distale :
    - A la différence des expériences propres du sujet, c'est la zone des normes et de la **culture**. Textuellement, c'est la zone marquée la troisième personne verbale, du récit, du non-présent et de l'irréel.

# 5. Sémantique Interprétative et TAL

---

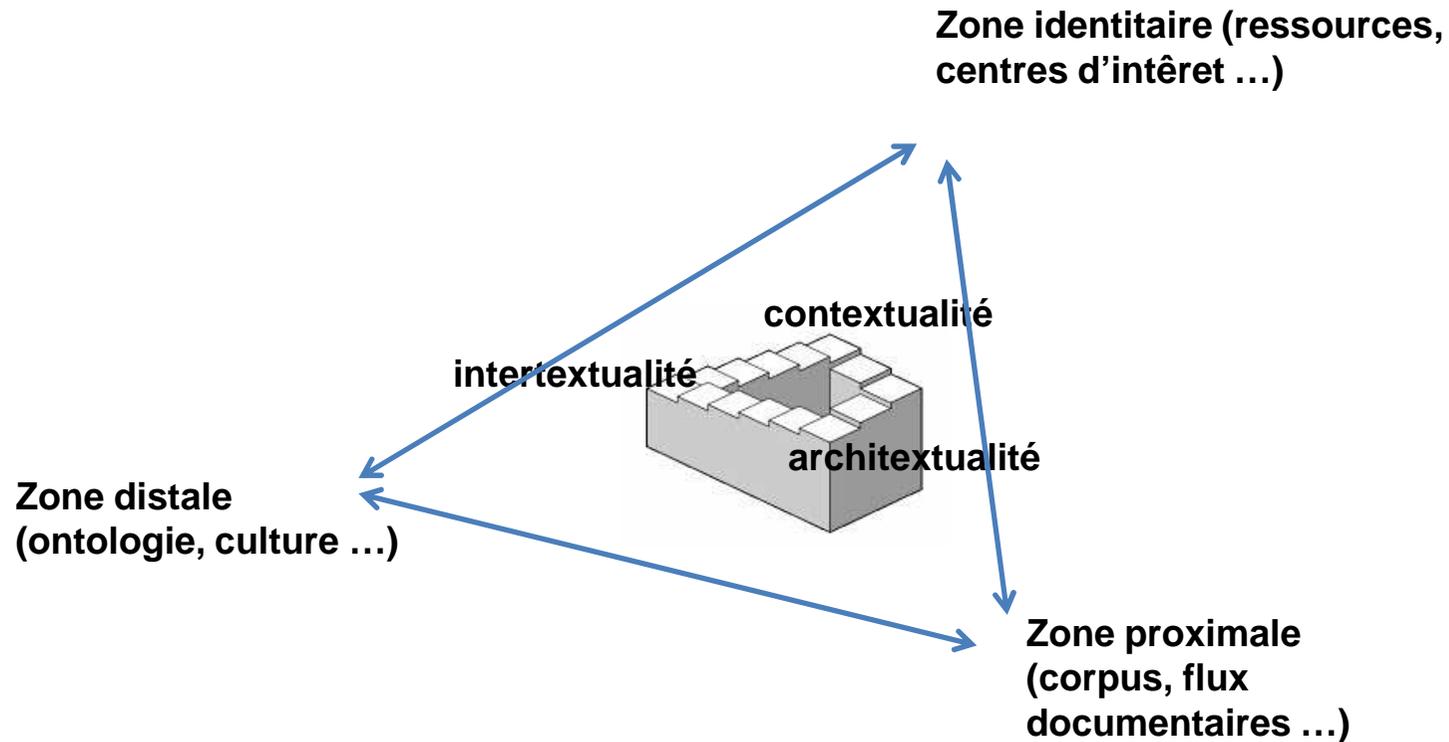
- Les niveaux de contexte d'interprétation :
  - Les zones sont des niveaux praxéologiques
  - Les phénomènes interprétatifs s'expliquent dans les rapports entre ces différents niveaux (*Interpréter, c'est s'y retrouver*)



# 5. Sémantique Interprétative et TAL

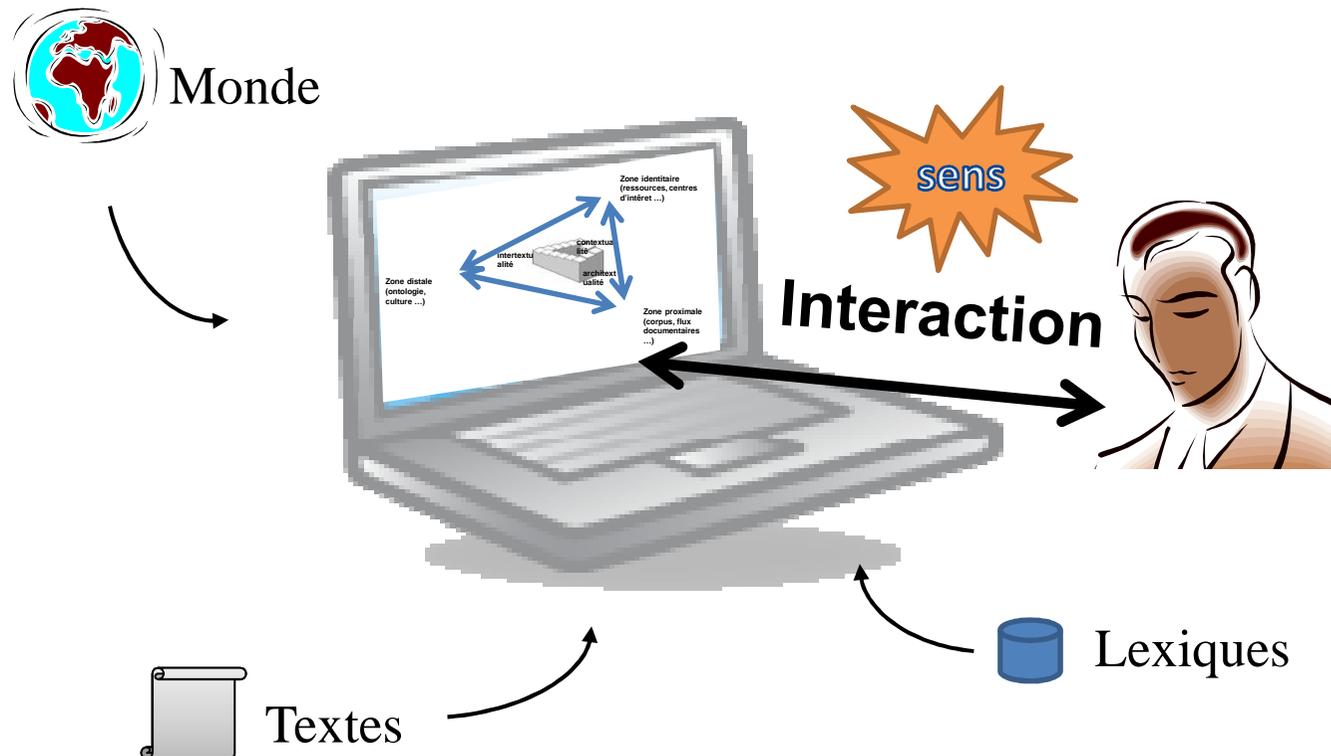
---

- L'espace interprétatif :



# 5. Sémantique Interprétative et TAL

- Médiatiser l'espace interprétatif



# Plan

---

1. Mon parcours
2. Objectifs et problématique
3. Travaux et développements
4. Expérimentations sur corpus
5. Sémantique Interprétative et TAL
- 6. Vers des ENT énatifs**
7. Conclusions

## 6. Vers des ENT éenactifs

---

- ENT : Environnement Numérique de Travail
  - Un constat : de plus en plus d'ENT, avec des utilisateurs interconnectés, des ressources documentaires, des flux de communication ...
    - Un exemple : les plateformes de e-learning Moodle (34 millions d'utilisateurs dans le monde en juin 2010)
- Quel cadre théorique pour développer des ENT ?
  - Un cadre où la subjectivité est première, où la boucle action/perception est centrale
  - Un cadre pour étudier les conditions d'un couplage Personne-Système :
    - les usages et les détournements/contournements d'usage

## 6. Vers des ENT énaectifs

---

L'énaction : théorie de **Francesco Varela** (1946-2001)

- Une théorie du vivant en biologie

Origine : l'autopoïese (Maturana & Varela – 1989)

L'invariant fondamental du système vivant est son organisation



Le concept d'**énaction** ou cognition incarnée permet d'appréhender l'action adaptative de tout organisme vivant comme polarité connaissance/action et action/connaissance.

L'énaction fait prédominer l'action et la boucle action ↔ perception à la représentation.

Exemple : Travaux de Paul Bach-y-Rita sur la suppléance perceptive

## 6. Vers des ENT énaactifs

---

- Une théorie qui peut être réinvestie dans d'autres champs de recherche que le vivant et induire quelques ruptures :
  - En systémique :
    - un système autonome n'est jamais un système isolé car il est nécessairement couplé à un milieu ambiant où il maintient son organisation en dépit des perturbations occasionnées par le milieu.
  - Pour les sciences cognitives :
    - une alternative au cognitivisme
    - un nouveau paradigme scientifique émergent non plus inspiré par la métaphore de l'ordinateur mais par celle du vivant.
  - Pour l'informatique et le TAL :
    - une remise en cause du statut des représentations
    - un tournant obligé vers la question du couplage personne/système, donc des interactions

## 6. Vers des ENT énaactifs

---

- La question de la représentation en TAL :

Remarque : une machine peut très bien effectuer des traitements sémantiques sans nécessairement manipuler des représentations du sens (méthodes statistiques et connexionnistes par exemple)

La question n'est pas de refuser les représentations mais d'en préciser le statut

Le statut des représentations relativement à l'énaaction :

**Une représentation est liée à une interaction avec un milieu**

↪ pas de représentations objectivées valables pour tout le monde quelle que soit l'activité (rupture par rapport au courant du W.S.)

↪ une représentation est **individuelle**, **située** car liée à une expérience du monde qui n'a pas de fin prévisible

## 2. Rapports entre TAL et Interactions

---

- La question de l'interaction et d'une instrumentation énaactive :
  - Le point de vue de Michel Beaudoin-Lafon (LRI-Orsay) :
    - Le courant des IHM s'est principalement consacré à développer des interfaces, mais finalement assez peu des interactions ...
    - L'interface idéale serait alors celle qui serait « invisible » laissant ainsi place à une interaction naturelle donnant notamment aux machines plus de capacités de résilience.
- Une instrumentation énaactive ne cherche pas à réduire une complexité pour mieux l'appréhender. La complexité est « dissoute » dans sa perception par l'outil produisant le couplage. L'outil énaactif n'est pas en soi perceptible ...



## 6. Vers des ENT éenactifs

---

- ENT éenactif : Le projet en cours de développement « Aidé »
  - Partenaires
    - Groupe NU (Nouveaux Usages) : un réseau normand « indisciplinaire » autoconstitué (LITIS, LiDiFra, GREYC, Modesco-MRSH Caen)
    - IDIT (Institut du droit international des transports), SAVE
    - Pôle Nov@log
  - Objectifs : En environnement qui
    - Permet d'interagir avec la base documentaire
    - Permet de prototyper et projeter sur corpus des ressources
    - Permet de tracer l'activité de l'utilisateur
    - Permet d'intégrer des outils en interopérabilité
    - Permet de faire émerger des usages non prescrits

## 6. Vers des ENT énéactifs

---

- « Aidé » : Veille dans le domaine du droit des transports (herméneutique juridique)
  - Interrogation par les adhérents de l'IDIT
    - Professionnels du transport
    - juristes, transporteurs, manutentionnaires, ports, intelligence économique ...
  - Corpus de l'IDIT :
    - 19571 fiches (documents secondaires interprétant des documents primaires de cour d'appel, cour de cassation ...)
    - Articles (17846), Textes officiels (2449), Fond documentaire (944), Guide des transports maritime et routier « destinés à tous ceux qui sont confrontés aux problèmes juridiques liés à une opération de transport de marchandises », Fiches *lexisnexis* du transport maritime et routier

# 6. Vers des ENT énonciatifs

- « Aidé » :

COMM.

M.F

COUR DE CASSATION

Audience publique du 18 avril 1989

M. BAUDOIN, président

Pourvoi n° P 87-14.850

Cassation partielle

Arrêt n° 564 P

Thèmes :

Avaries par mouille  
responsabilité du transporteur Colis conteneur

Date de la décision : 18/04/1989

Mode de transport : Maritime

Pays : France

Objet :

Avarie par mouille (sac de semence de tournesol) - Responsabilité du transitaire (non).

Sommaire :

Le transitaire chargé de procéder à la réception des marchandises n'est tenu de réserver le recours de son mandant qu'en présence de dommages extérieurs apparents ou de dommages que les diligences normales auraient permis de découvrir. En présence de dommages non apparents est considéré comme étant des diligences normales, le fait pour un transitaire d'avoir formulé des réserves, et provoquer une expertise. Lorsqu'un conteneur est utilisé, pour le transport de colis ou d'unités énumérés au connaissement, ce sont les colis ou unités qui sont pris en compte pour déterminer le montant de l'indemnité (article 1er du décret du 23 mars 1967). Ainsi la clause qui prévoit que l'indemnité sera calculée en fonction du nombre de conteneur est réputée nulle.

- Référence :

Cour de cassation (Ch. Com.)

18 avril 1989

Pourvoi n° P 87-14.850

Royal Globe Insurance Company LTD c/. Dart Container Line campagny limited of Reid house et autres

Bulletin des Transports - 1989 - p. 498.

Droit Européen des Transports 1990 p. 646

- Observation :

- Décision intégrale disponible à l'IDIT

R E P U B L I Q U E F R A N C A I S E

A U N O M D U P E U P L E F R A N C A I S

LA COUR DE CASSATION, CHAMBRE COMMERCIALE, a rendu l'arrêt suivant :

Sur le pourvoi formé par LE ROYAL GLOBE INSURANCE COMPANY LTD, représentée par la Société LANDIRE ET CHEVILLAT, dont le siège social est 19, boulevard Montmartre à Paris (2ème),

en cassation d'un arrêt rendu le 9 avril 1987 par la cour d'appel de Rouen (2ème chambre), au profit :

1°/ de la Société DART CONTAINER LINE COMPANY LIMITED OF REID HOUSE, dont le siège social est Hamilton Bermudes,

2°/ de la Société DART CONTAINER LINE LIMITED, dont le siège social est 1600 Darchester, BD West Montréal H 3 H 1 R 2, représentée par la Compagnie Maritime des Chargeurs Maritimes agence du Havre boulevard de Strasbourg,

vu l'article 1er de l'article 1er du décret du 23 mars 1967 ;

Attendu qu'il résulte des textes susvisés que, dans le cas où un conteneur est utilisé pour le transport de colis ou unités énumérés au connaissement, ce sont ces colis ou unités, qui sont pris en considération pour déterminer le montant de l'indemnité prévue à l'article 28 de la loi du 18 juin 1966, et destinée à réparer des pertes ou avaries, toute clause contractuelle ayant pour objet ou effet de limiter à une somme inférieure la réparation due par le transporteur maritime étant réputée nulle ;

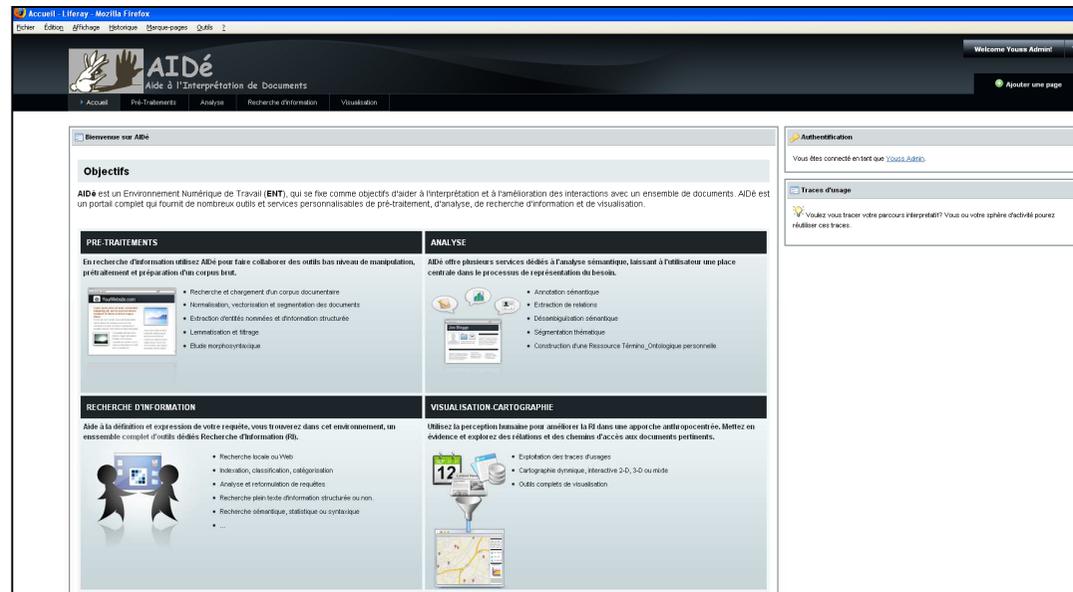
Attendu que, pour décider que le montant de l'indemnité due en raison des avaries devait être calculé par rapport au nombre de conteneurs, et non en fonction de celui des sacs de semences avariés dont le fait avait constaté que la description et le numérotage figuraient sur le connaissement, la cour d'appel a retenu qu'une clause du connaissement définissait l'unité "comme étant le conteneur" ; qu'en statuant comme elle a fait, la cour d'appel a violé les textes susvisés ;

Fiche IDIT

# 6. Vers des ENT éenactifs

- « Aidé » : pistes pour le développement
  - Intégrations d'outils sur les différentes tâches de l'utilisateur dans l'accès au contenu (statistiques textuelle, OCR, indexation ...)
  - Intégrations d'outils anthropocentrés ou anthropocentrables

- LUCIA,  
ProxiDocs,  
Nooj ,  
Activa ...



## 6. Vers des ENT énaactifs

---

- « AIDé »
  - Rendre l'utilisateur « cré-actif »
  - Faire émerger les problèmes plus que des « solutions »
  - Une recherche d'information n'est pas une extraction d'information, mais plutôt une recontextualisation de l'information, une retextualisation de l'information
  - le système cognitif de l'utilisateur interagit avec son ENT en subissant des perturbations qui l'incitent à évoluer dans sa recherche d'information ...
  - Une interaction (non nécessairement finalisée) plus qu'un résultat (le « couplage personne-système » plus que les « entrées-sorties »)

## 6. Vers des ENT éenactifs

---

- Evaluer un ENT éenactif ?
  - Pour évaluer il faut a priori avoir un attendu de résultat et c'est contraire au principe de l'éenaction
  - Comment anticiper une appropriation, un contournement d'usage ?
  - Une interprétation n'est pas « reproductible à l'identique » (mais les conditions d'interprétations, oui ...)
  - Un ENT éenactif, ça ne s'évalue pas, ça s'expérimente
  - Protocoles expérimentaux et entretiens d'explicitation

# Plan

---

1. Mon parcours
2. Objectifs et problématique
3. Travaux et développements
4. Expérimentations sur corpus
5. Sémantique Interprétative et TAL
6. Vers des ENT énatifs
7. **Conclusions**

# 7. Conclusions

---

- Quelles difficultés
  - Les « abandonware »
  - Le cloisonnement disciplinaire et les clivages au sein des disciplines (ex : TAL & IHM)
  - Le passage à l'échelle (plus de documents, plus d'utilisateurs, plus de ressources, plus d'usages ...)
  - Les protocoles expérimentaux et les évaluations
  - ...
- Quelles « chances » pour les approches anthropocentrées
  - L'utilisateur qui ne demande qu'à interpréter et faire émerger du sens
  - Des outils simples qui induisent le couplage ...
    - ... du couplage découle une prise en compte de la complexité

# 7. Conclusions

---

- Il ne s'agit pas de « calculer du sens » (Il n'y a pas de fond sémantique sans forme, pas de formes extralinguistiques du sens).
- Il s'agit de mettre en place des interactions homme-machine qui **font sens**.
- Le sens ne connaît pas d'état final « stable ». C'est une activité.
- La façon dont on cherche à accéder au contenu, c'est déjà le contenu.
- Les interprétations sont « observables » a posteriori mais pas prédictibles.
- Le sens est autant le fait de son lecteur que celui de son auteur. Il convient de subjectiver les processus et ressources (comprendre, c'est comprendre autrement).
- Une instrumentation sémantique n'est pas forcément un modèle prédictif. Cela peut-être une interaction sémiotiquement suggestive.