

# Nathalie Aussenac-Gilles

Institut de Recherche en Informatique de  
Toulouse

45 ans  
3 enfants  
ENSEEIH  
HDR 2005

CR CNRS depuis 1991 en Intelligence artificielle

Responsable de l'équipe IC3 :

ingénierie des connaissances, de la cognition et de la coopération

Modélisation des connaissances pour et par l'accès  
au contenu textuel

Problématique

Contribution

Projet scientifique

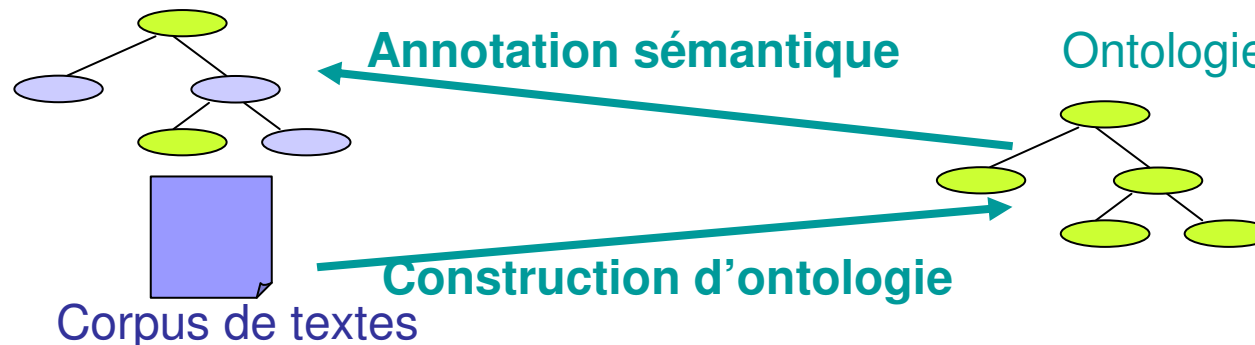
Bilan



# Ontologies et textes : enjeux



- **Modéliser les connaissances POUR l'accès aux contenus textuels**
  - Importance des besoins applicatifs
    - Accès au contenu de collections documentaires numérisées, gestion patrimoniale
    - Systèmes à base de connaissances
  - Approche Web Sémantique
    - Formaliser les contenus (associer des annotations sémantiques)
    - S'appuyer sur des ontologies



- **Modéliser les connaissances PAR l'accès aux contenus textuels**
  - Construire les ontologies à partir des textes
  - Diversifier les sources de connaissances ontologiques

# Ontologies et textes : enjeux



- **Problématiques de recherche**

- Difficultés théoriques
  - Multiplicité des interprétations, ambiguïté de la langue
  - Diversité des connaissances

“*American General Electric CF6-50 engines powered the A300*”

<CF6-50 engines> POWER <A300>

(relation entre instances)

<engineModel> POWER <planeModel>

(relation entre propriétés de concepts)

<engine> IS-PART-OF <plane>

(relation entre concepts)

- Difficultés pratiques
  - Fixer une sémantique, des cadres méthodologiques et des outils
  - Fixer les interprétations des textes en fonction des objectifs applicatifs
  - Valider les modèles : formellement, par des méta-propriétés, par l’usage

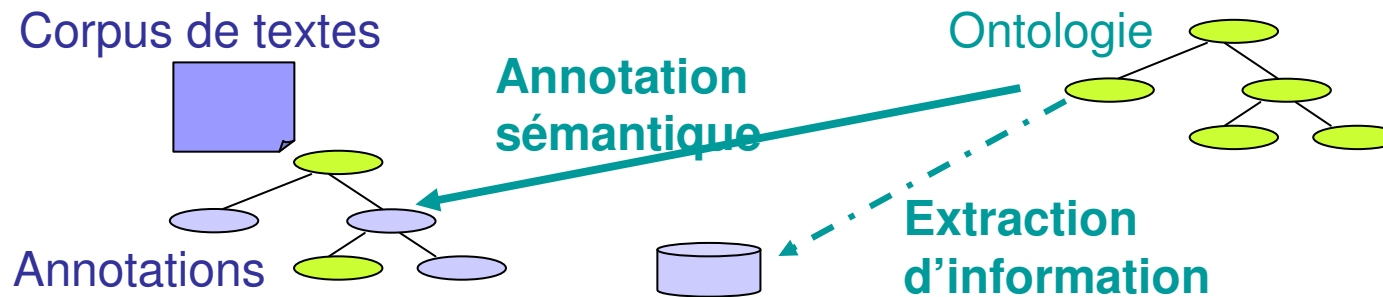
# Des textes vers les ontologies : contributions



**Construire une ontologie, c'est décider d'un point de vue sur un domaine, fixer une interprétation normalisée d'éléments pertinents pour l'application**

- Quelle sémantique ?
  - *sémantique textuelle (ERSS linguistique) (Applied Ontology J.)*
- Quelle méthode ?
  - *Approche linguistique à partir de textes (Terminae, LIPN) -> projet ANR (EKAW RIA)*
- Quelles techniques d'analyse de textes ?
  - *Extracteur de terme (évaluation de Syntex – ERSS)*
  - *Extracteur de relations Caméléon (Th. Séguéla – ERSS) (Terminology, TAL)*
  - *Système Multi-Agents adaptatifs (Th. Ottens, Th. Sellami – IRIT-SMAC) (JLC)*
- Quelles représentations ?
  - *Associer une composante lexicale à une représentation logique (Th. Reymonet) (OntoLex)*
- **Constat** : traitement de la langue découplé de l'édition d'ontologies

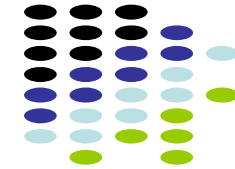
# Des ontologies vers les textes : contributions



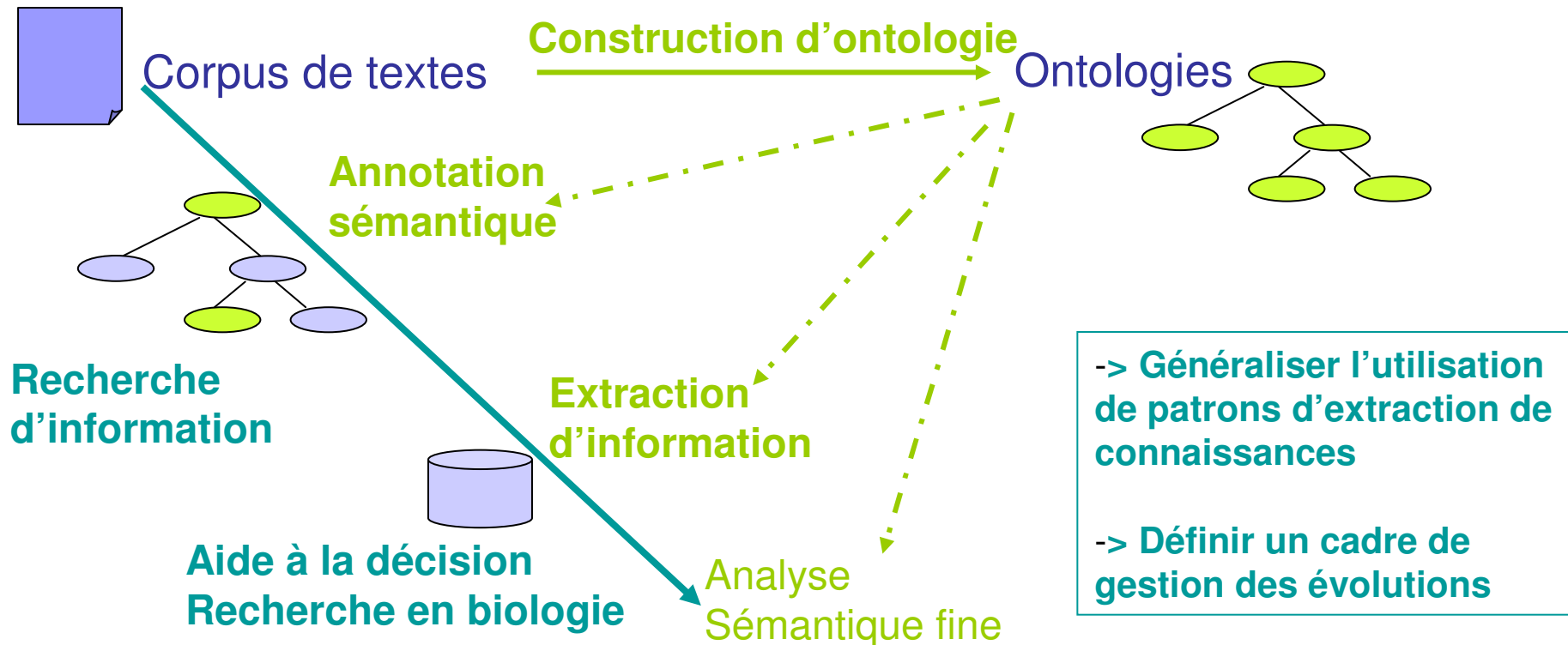
***Annoter sémantiquement, c'est projeter sur les textes une interprétation, celle fixée par l'ontologie, dont le contenu doit être pertinent pour l'application***

- Comment identifier des éléments d'ontologie dans les textes ?
  - *Désambiguïsation des termes (IRIT-SIG, Th. Baziz)*
- Quelles représentations des textes pour la recherche d'information ?
  - *Graphes de concepts pondérés / graphes d'instances / concepts (CLEF, TREC, COLIS)*
- Comment interroger les annotations ?
  - *Formulation et expansion de requêtes ; distances sémantiques (Th. Reymonet) (SWIR)*
  - *Navigation via l'ontologie, clustering de collections documentaires (Th. Hernandez)*
- **Constat** : faible exploitation du traitement automatique des langues

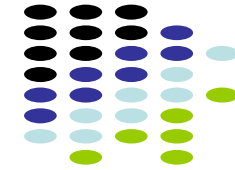
# Projet de Recherche



- **Intégrer analyse de la langue et édition d'ontologies**
- **Penser en cohérence construction et utilisation des ontologies**
- **Prendre en compte la dynamique des collectifs d'utilisateurs**
- **Diversifier les applications cibles (aide à la décision)**



# Généraliser l'utilisation de patrons d'extraction de connaissances

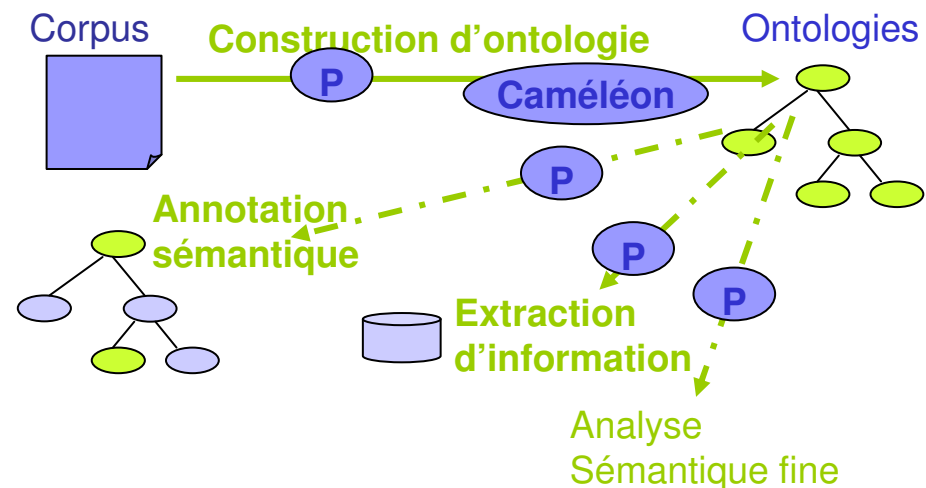


“*American General Electric CF6-50 engines powered the A300*”

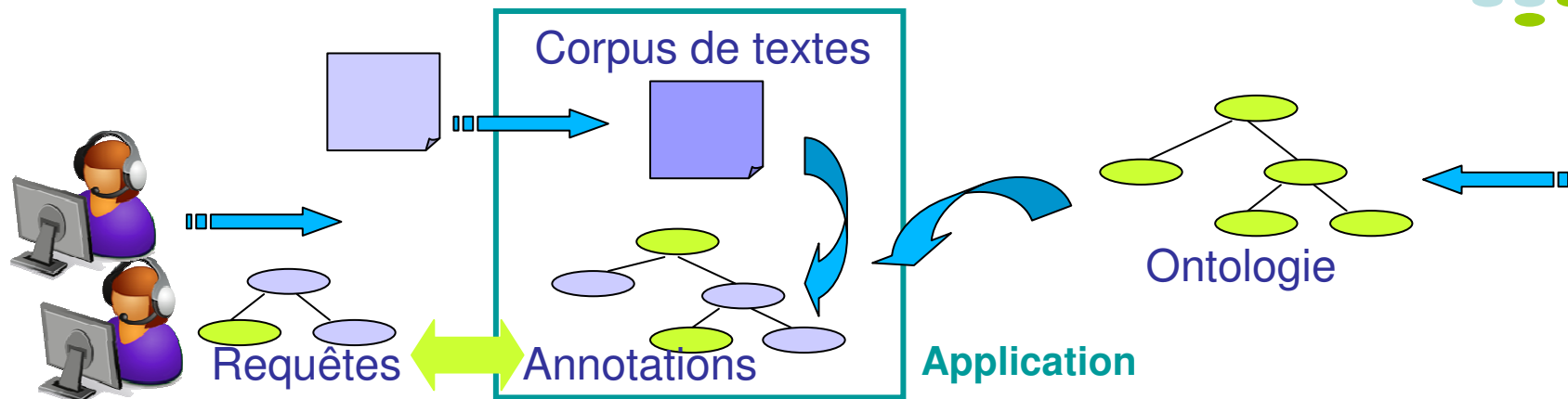
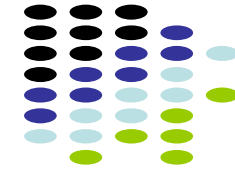
<engineModel>	to power + the	<planeModel>
NP	to power + det	term2
<engineModel>	verb + det.	<planeModel>

- **Patron** Caractérisation des régularités de sens par des régularités de forme linguistique
- **Etendre la notion de patron**
  - Repérer des concepts, des règles
  - Repérer des relations autres que binaires, relations sur plusieurs phrases (TAT-CG)
- **Homogénéiser les patrons pour différents usages**
  - Construction / annotation / extraction
  - Capitaliser des patrons

-> logiciel unique,  
compatible avec plates-  
formes existantes,  
diffusable



# Vers un cadre de gestion des évolutions



- **Environnements fortement dynamiques**
  - Situations collectives - Contextes évolutifs
  - Annotations, documents, ontologies, besoins utilisateurs
- **Assurer l'équilibre de l'ensemble**
  - Définir les évolutions et leurs conséquences (Dynamo, Th Tissaoui)
  - Définir des critères de stabilité, de cohérence
  - Système supervisant la gestion de la cohérence
  - Dimension ergonomique : évolution des besoins des utilisateurs
- **Ontologies comme aide au repérage d'évolutions dans un domaine**

# Responsabilité de l'équipe IC3

Ingénierie des Connaissances, de la Cognition et de la Coopération



- **Thèmes de recherche** (13 permanents, 14 doctorants et contractuels)
  - Ontologies et textes (8)
  - Coopération et décision multi-critères (3)
  - Coopération au sein des situations collectives et des systèmes complexes (2)
- **Collaborations au sein de l'IRIT**
  - Recherche d'information et ontologies : plate-forme ExPRIM, SIG-EVI (2 thèses)
  - Systèmes multi-agents pour la construction d'ontologies : SMAC
  - Traitement automatique des langues : LiLAC, groupe « langage naturel »
  - **Axes prioritaires** “masse de données” et “santé” de l'IRIT
  - **Membre de conseils scientifiques** : IRIT, UFR MIG de l'UPS, dept. Info de l'UTM
- **Encadrements**
  - **7 thèses terminées** (co-encadrements avec SIG, SMAC, ERSS)
  - **4 thèses en cours** (co-encadrements avec LaLIC, SMAC et ERSS)
  - 8 M2R et près de 30 stages ingénieur
- **Participation à des enseignements**
  - **M2R** ECIL, MITT (resp module ALRI), ICMST, ENSEEIHT, URFIST, UTT (25h /an)
  - **Membre de commissions de spécialistes** en 27<sup>e</sup> : UTM (2007-2008) et UPS (1996-1997)
  - Organisation de 15 séminaires sur le traitement automatique des langues ERSS-IRIT

# Collaborations, animations nationales



- **Animations nationales**
  - Groupes de travail : GRACQ (20 journées de travail), thème 7 du GDR-I3, “Terminologie et IA”
  - AS du RTP-DOC “Terminologie et Corpus”
  - Comité de pilotage des conférences IC et TIA, CP de 8 conférences
  - Organisation de conférences : IC 2000, TIA 2009, “semaine de la connaissance 2006”, 2 ateliers
  - Comité de rédaction de la Revue I3
- **Collaborations nationales**
  - Méthodes de modélisation : LRI, IRIN, INRIA-Acacia, TechCico, DER-EDF, LaRIA
  - Construction d’ontologie à partir de textes : CLLE-ERSS, LIPN, INSERM-S\_852, INRIA-Edelweis
  - Extraction de relations sémantiques : CLLE-ERSS, LIPN, LIUPPA, COGIT
  - Annotation sémantique : “Préhistoire et Technologie”, LaLIC, Mondeca
- **Projets et contrats**
  - Régionaux (4), nationaux (9) :TCAN, Sciences de la cognition
  - 4 projets ANR : Dynamo, DAFOE4App, Geonto, Corpus logicistes
  - Contrats avec entreprises (11) : EADS, CNES, CEA, DER-EDF, Saint-Gobain
  - Laboratoires mixtes : Aramihs (Matra), Autodiag (ACTIA, LAAS-Disco)
- **Evaluations**
  - Participation à des jurys de thèse (14) et HDR (2)
  - Comité d’évaluation du programme ANR-DEFI, évaluations RNTL, Masse de données, ANVAR
  - Membre nommé de la CID 42 du comité national du CNRS (2008-2012)

# Collaborations, animations internationales



- **Animation internationale**

- Resp. thème “lexique et ontologies” du LEA ILIKS
- Comité de pilotage de la conférence EKAW (+ organisation de la 4<sup>e</sup> en 1993)
- Comité éditorial des revues : Applied Ontology et IJ Human Comp. Studies
- Co-organisation de workshops : Ontologies & Texts (3), Indexation et Connaissances en SHS
- Comité de lecture de 5 revues, CP de 10 conférences et 20 workshops

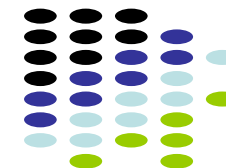
- **Collaborations internationales**

- Ingénierie des connaissances : N. Tourigny (Univ. Laval), J. Breuker (Univ. Amsterdam)
- Ontologies et textes : IIC -Univ. Murcia, LOA – univ. Trento, DFKI, Karlsruhe (T. Declerk, P. Buitelaar), AIFB (P. Cimiano)

- **Séjours à l'étranger**

- LOA, Univ. de Trento – 2007 (ILIKS)
- LBMI, Stanford Univ. – 2008
- Univ. Murcia (professeur invité) – 2005, 2008, 2009

# Publications



Type	National	International
Revue	10	5
Chapitres d'ouvrage	12	2
Conférences actes édités	21	18
Ateliers	12	17
Edition	1 livre, 2 n° revue, 4 actes	3 actes

- **Notoriété des revues** IJ Human Computer Studies,  
Applied Ontology, Journal of Logic and Computation, Terminology
- **Principales conférences** et challenges du domaine
  - EKAW, KES, SAC, COLIS,
  - TREC, CLEF,
  - workshop OntoLex et Ontology Learning and Population...
- **Logiciels** : Macao, Caméléon, Géditerm, prototype Arkeotek