

# ATOLL - INRIA Rocquencourt

<http://atoll.inria.fr>

## Petit état des lieux

Éric de la Clergerie

[Eric.De\\_La\\_Clergerie@inria.fr](mailto:Eric.De_La_Clergerie@inria.fr)

Réunion RNIL

AFNOR – 23 Septembre 2003

## Réunions récentes et à venir

- Réunion TC37 SC4 (Sapporo, Juillet 2003, N062)
- Réunion TC37 (Oslo, Août 2003)
  
- Réunion jointe TC37SC4/TEI sur les Structures de Traits (Nancy, 5-8 Novembre 2003)
- Morpho-Syntaxe (Fin Octobre/ Début Novembre 2003)
- Réunion Corée Février 2004 : Structures de traits et Morpho-Syntaxe
- SC4 à LREC (Lisbonne, Mai 2004)
- Mi-Août SC4TC37 en France

## Chantiers en cours

- Structures de Traits (**FSR** & **FSD**, Kiyong Lee, Corée)
- Morpho-Syntaxe (**MSAF**, Lionel Clément, Eric de la Clergerie)  
Morpho-Syntactic Annotation Framework
- Catégorie de Données (**DCR**, Laurent Romary)
- Cadre d'Annotation Linguistique (**Linguistic Annotation Framework** [LAF], Nancy Ide)
- Propositions de nouveaux domaines de normalisation

# Structures de traits

- Draft **N033** (Février 2003)  
Plus spécifique Représentation Structures de Traits (FSR)  
Manque Declaration de Structures de Traits (FSD) ~ DTD
- Votes (Juin 2003, **N051**) :  
NWI (Nouvel Item de Travail) : 1000%  
Committee Draft (CD) Ballot : presque 100% (mais peu de commentaires)
- Préparation de commentaires français  
À discuter au sein de RNIL et sur la liste RNIL
- Réunion jointe SC4/TEI sur FS en Novembre à Nancy  
Meilleure coordination avec TEI  
Accord TEI d'intégration de propositions
- Préparation d'un draft sur FSD pour Février 2004
- Réunion en Corée (Février 2004) sur FSR & FSD

# Mopho-Syntaxe

- Votes (SC4 - Oslo)  
NWI : 100%
- Draft pour Octobre 2003
- Réunion technique, AFNOR courant Novembre ?
- Draft révisé pour Décembre 2003
- Réunion jointe en Corée (avec FSR/FSD) en Février 2004
- Document pour CD Ballot, Février 2004
- Décision après vote lors de LREC (Lisbon, Mai 2004)

# Catégories de données

- Présentation à Sapporo (Laurent Romary, N065)
- Discussions à Oslo
- Implantation d'un repertoire de catégorie de données pour TC37SC4  
Constitution d'un fond pour la morpho-syntaxe
- Réflexions : faut-il évoluer vers les ontologies ?

# Linguistic Annotation Framework – LAF

- Identification de formalismes génériques et pivots en Linguistique
- Séminaire Pont-à-Mousson (Novembre 2002, N029)
- Séminaire ACL (Sapporo, Juillet 2003)
- Proposition de NWI (N076 + N029 & N079)  
Vote pour 15 Novembre 2003

Recognizing the diverse needs for representing different types of linguistic annotation, divergences in theoretical approach, and the existence of large bodies of legacy data and software, the proposed project does not seek to establish a single, definitive annotation scheme or format. Rather, the goal is to provide a framework for linguistic annotation of language resources that can serve as a reference or pivot for different annotation schemes, and which will enable their merging and/or comparison.

## LAF – survol

Une infrastructure générique pour les ressources linguistiques et des mécanismes pivots.

- Pouvoir d'expression linguistique ; indépendance aux media ; association d'une sémantique.
- Incrémentalité ; gestion sous-spécification et alternative ; uniformité (réutilisation de « modules »)
- ouverture et neutralité ; mécanismes d'extension (DCR) ; lisibilité humaine
- Modèle : Graphes de FS associées à des régions (points, intervalles, ...) primaires ou à des annotations de niveaux supérieurs.
- Format Pivot (XML) : basé sur GMT (Ide et Romary)  
structures ; alternatives ; conjonction ; FS

```
<struct type="W-level">  
  <seg target="#w3">  
    <feat type="lemma">de</feat>  
    <feat type="pos">PREP</feat>  
  </struct>
```

```
<W-level xi:type="struct"  
  xlink:href="#w3">  
  <lemma xsi:type="feat">de</lemma>  
  <pos xsi:type="feat">PREP</pos>  
</W-level>
```



# Propositions de nouveaux chantiers

- Multimodal Content Representation
- Reference Annotation Scheme
- Lexiques
- Mémoires de traduction

Notes : Contributeurs bienvenus !

# Scénario de déploiement

- Exploiter les divers axes en cours
- ⇒ Retour et validation des divers choix
- Exploitation des Répertoires de données en Morpho-Syntaxe  
Constitution d'un fond morpho-syntaxe  
Utilisation dans un ou plusieurs Tagset pour le Français
- Production de corpus annotés morpho-syntaxe dans les divers variantes proposés  
⇒ outils, convertisseurs, visualisation
- Plongement dans LAF  
Essais de conversion vers dump format (GMT étendu ?)
- Recherche de cobayes !