

*Fondements d'un prototype pour la création  
semi-automatisée de productions visuelles pour le génie  
logiciel à partir de documents corporatifs*

Pierre André Ménard  
Sylvie Ratté

Laboratoire d'Ingénierie Cognitive et Sémantique (Lincs)

École de Technologie Supérieure  
Montréal

# [ Plan de la présentation ]

- Contexte d'application
- Buts, exigences et contraintes
- Prototype
- Prochaines étapes
- Questions

# Contexte d'application

- Une entreprise qui engage un  
Projet logiciel demande l'intervention d'ingénieur logiciel
- L'analyse du domaine d'affaires est la première phase  
Pour mieux comprendre le domaine de connaissance et le contexte du projet
- Les entreprises matures possèdent beaucoup de documentation sur leur fonctionnement interne  
Doc interne sur les processus , cmmi, iso, ieee
- Cette documentation est une source de connaissances de choix lors de l'analyse de projets logiciels
- Le génie logiciel est peu préparé pour exploiter efficacement cette source

# [ Buts du projet ]

- Traiter une documentation corporative composée de multiples documents de différents types

De quelques centaines de pages à plusieurs milliers, selon la taille de l'entreprise et la couverture du projet

- Représenter les connaissances dans un format propre au génie logiciel

- Exécuter le traitement complet de façon semi-automatique

-Prototype ne nécessite peu ou pas d'intervention  
-Traitement amélioré par l'intervention humaine  
-Produit final doit être remanié

# [ Exigences ]

- Traiter les différents types de documentation d'une entreprise mature
- Ne pas recourir à des experts du domaine
- Traiter le français naturel écrit
  - Sans passer par réécriture langage formel
- Conserver tous les concepts pertinents
  - Quitte à avoir un peu de bruit, sinon inutile, aide à la saisie
- Permettre la traçabilité des informations
  - Remonter de la représentation finale vers les textes source

# [ Contraintes ]

- Internes Imposées pour restreindre la complexité du projet
  - Traiter uniquement les documents électroniques
  - Produire un modèle du domaine avec notation UML 2
- Externes
  - Faible disponibilité des ressources linguistiques francophones adaptées

Adaptés aux traitements automatiques informatiques

# [ Exemples ]

- Phrase simple  
« Le registraire examine chaque demande d'admission. »
- Représentation visuelle UML 2

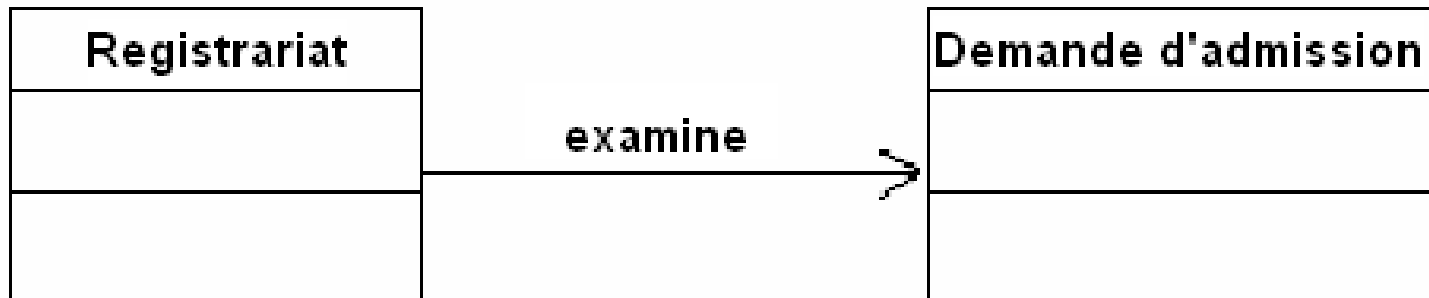
# [ Exemples ]

- Phrase simple  
« Le **registraire** *examine* chaque **demande d'admission.** »
- Représentation visuelle UML 2



# Exemples

- Phrase simple  
« Le **registraire** *examine* chaque **demande d'admission**. »
- Représentation visuelle UML 2



# Exemples

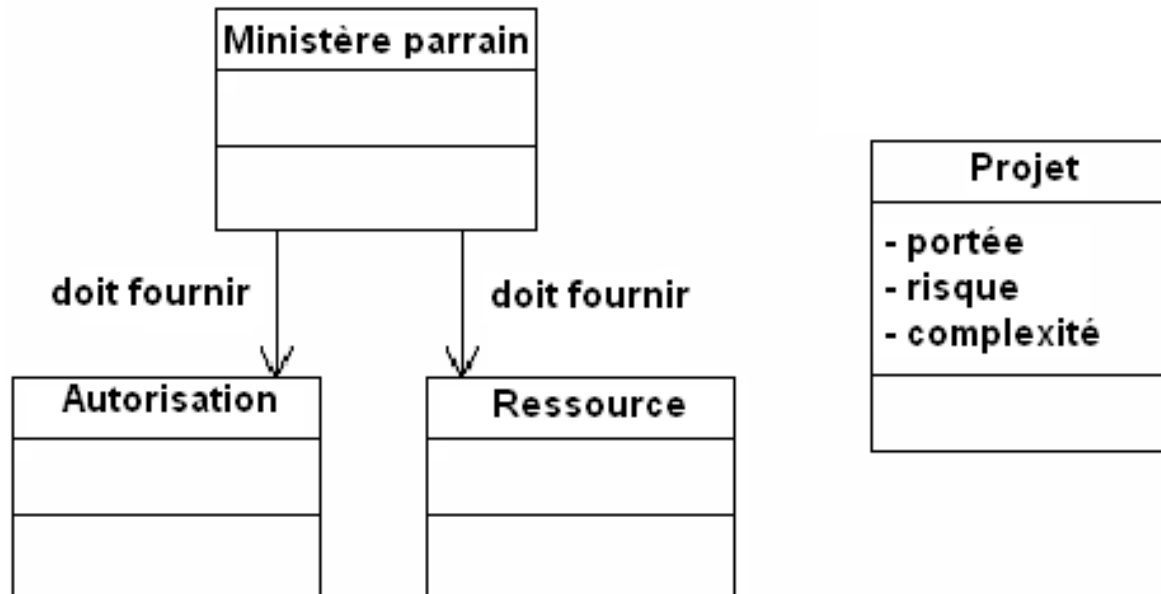
- Phrase plus complexe
  - « Les ministères parrains doivent fournir les autorisations et les ressources nécessaires, compte tenu de la portée, des risques et de la complexité du projet.»
- Représentation visuelle UML 2

# Exemples

- Phrase plus complexe
- « Les **ministères parrains** *doivent fournir* les **autorisations** et les **ressources** nécessaires, compte tenu de la **portée**, des **risques** et de la **complexité** du **projet**.»
- Représentation visuelle UML 2

# Exemples

- Phrase plus complexe
- « Les **ministères parrains** *doivent fournir* les **autorisations** et les **ressources** nécessaires, compte tenu de la **portée**, des **risques** et de la **complexité** du **projet**.»
- Représentation visuelle UML 2

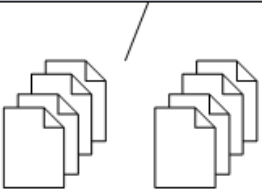


# [ Prototype ]

- Version en cours de développement
  - Extraire uniquement les concepts et leurs attributs  
**100% + bruit**
- Stratégies
  - Identifier tous les termes candidats
  - Utiliser des techniques pour éliminer le bruit  
**Expliquer le bruit**
  - Repousser l'usage de ressources complexes  
**Analyse syntaxique complète, lexique spécialisé en gl., etc.**

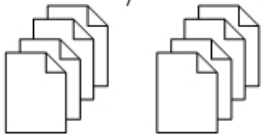
# [ Prototype ]

- Documents catégorisés
- Orthographe valide



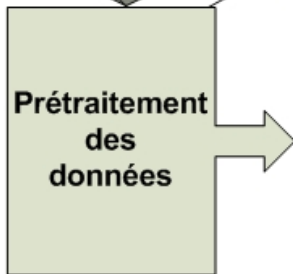
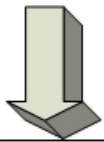
# [ Prototype ]

- Documents catégorisés
- Orthographe valide

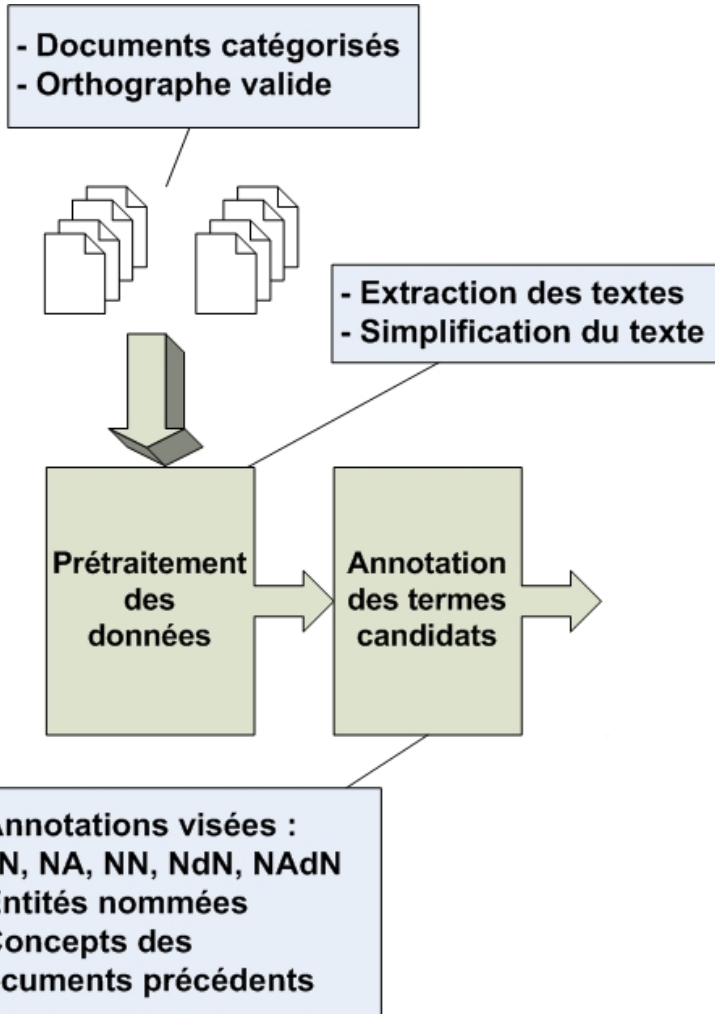


- Extraction des textes
- Simplification du texte

- Différents formats de document (pdf, word, etc..)
- Expressions vides ou simplement de forme

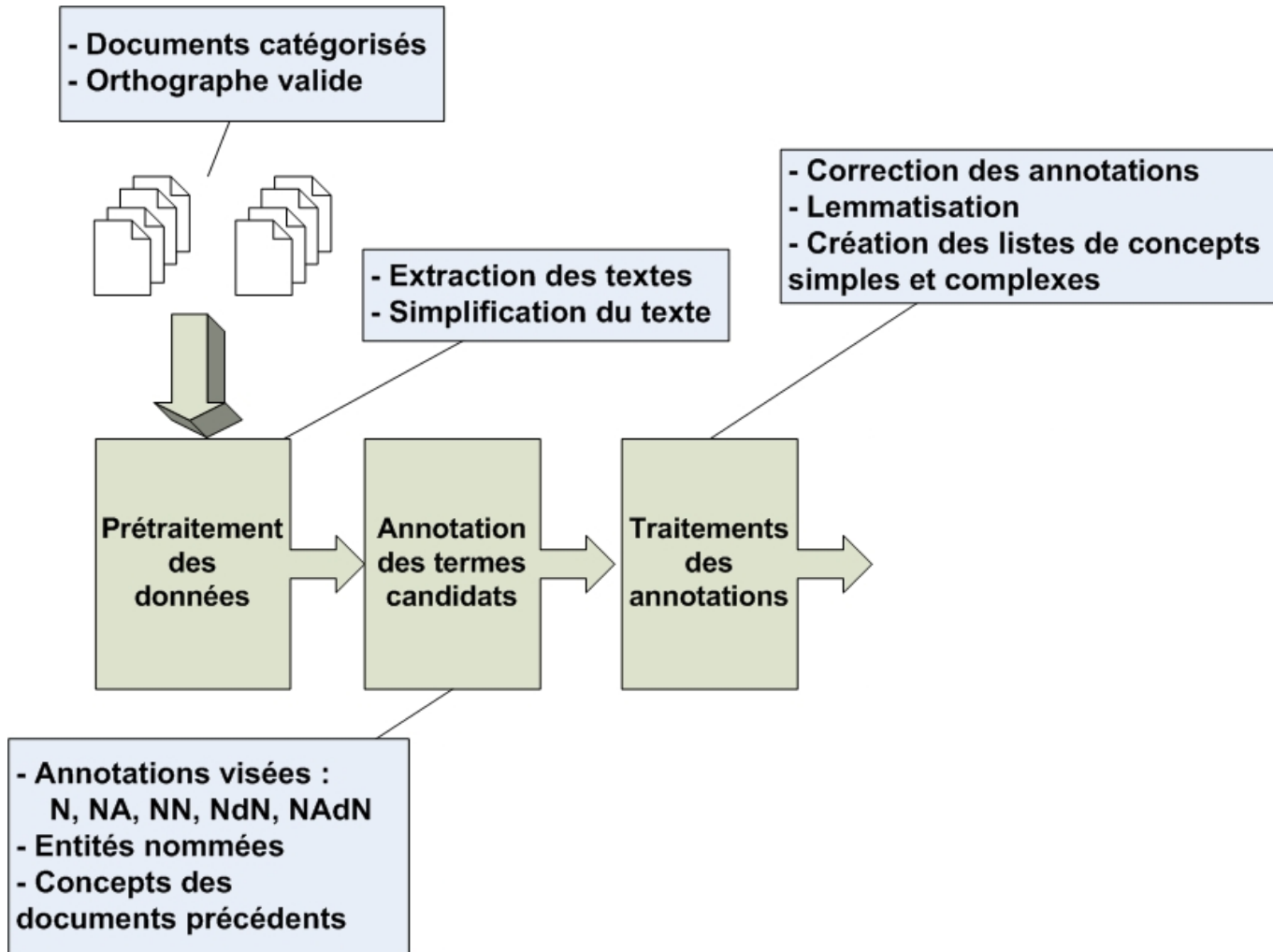


# [ Prototype ]

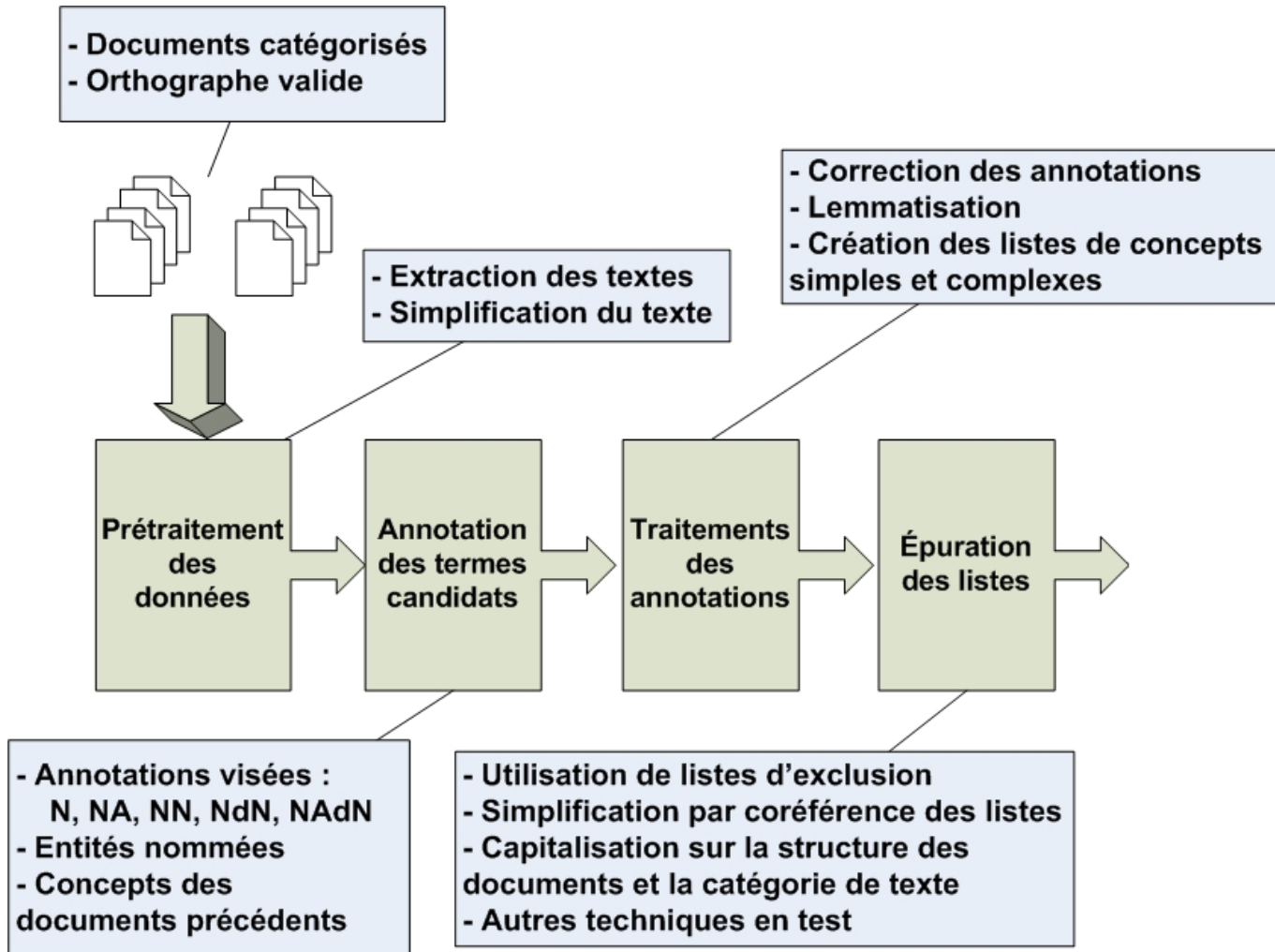




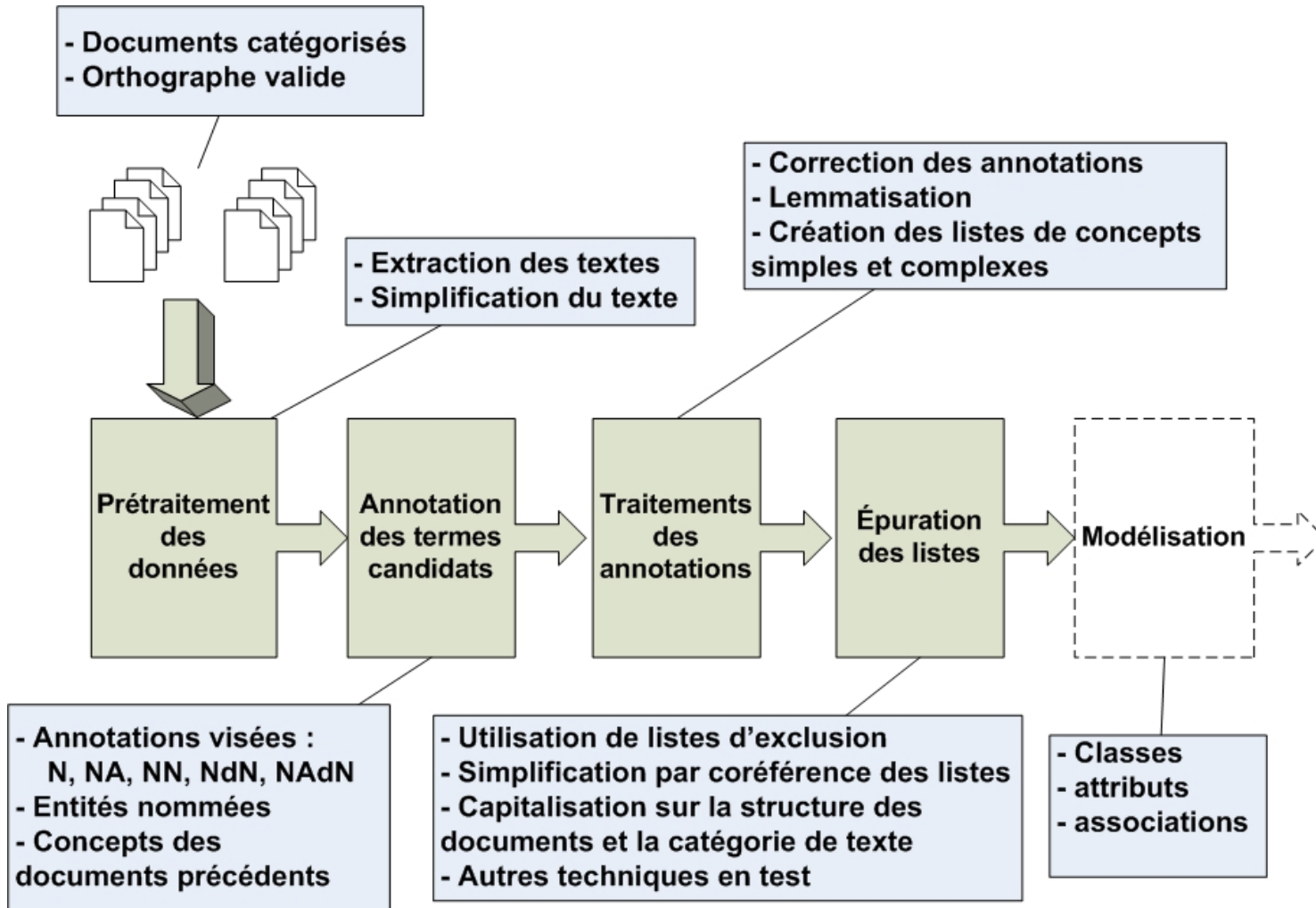
# [ Prototype ]



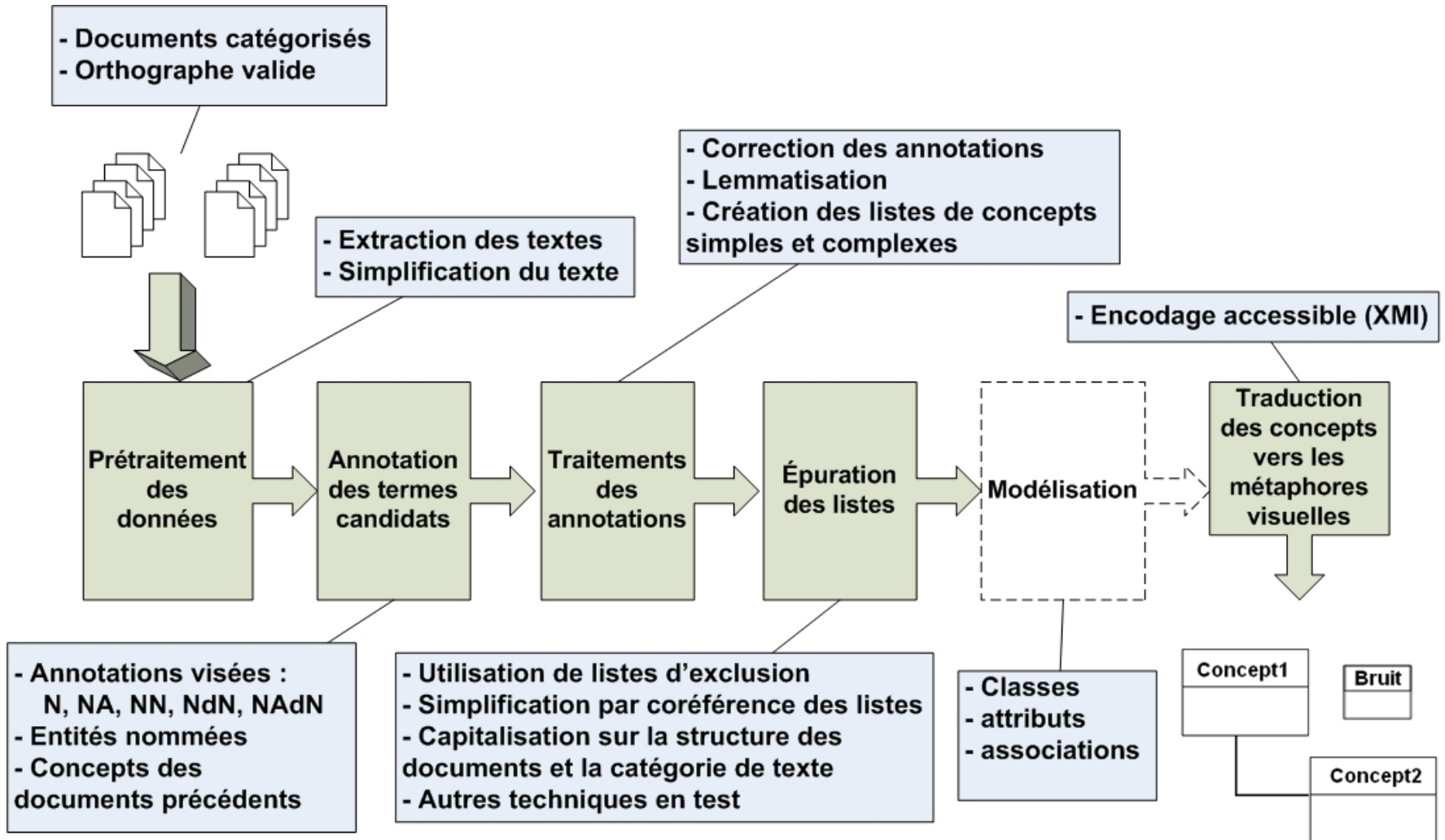
# [ Prototype ]



# [ Prototype ]



# [ Prototype ]



# [ Prochaines étapes ]

- Phases à venir
  - Ajout des liens entre les concepts
  - Modélisation des classes et attributs
  - Ajout des méthodes dans les concepts
- Autres possibilités d'enrichissement
  - Cardinalité des liens
  - Règles d'affaires
  - Autres types de modélisation

# [ Contributions ]

---

Organismes et entreprises désirant aider la recherche en contribuant leur documentation électronique.

Contactez-nous aux coordonnées suivantes

Site web : **<http://lincs.etsmtl.ca>**

Courriel : **[pamenard@gmail.com](mailto:pamenard@gmail.com)**

# [ Questions ]

---



Site web : <http://lincs.etsmtl.ca>  
Courriel : [pamenard@gmail.com](mailto:pamenard@gmail.com)

[

]



# [ Méthodes manuelles ]

- Peu de méthodes manuelles en GL
  - Abbot
  - Shlaer-Mellor
  - Coad et Yourdon
- Certaines méthodes sont trop floues :
  - Entités tangibles, rôles, incidents, spécifications, etc.
- Ces méthodes ne sont incomplètes et imprécises
- Elles sont adaptées principalement pour l'anglais

# [ Systèmes similaires ]

- En anglais

- Visual Paradigm **Des outils d'aide à la saisie, sans analyse auto**
- Lida de Cogentex

- En français :

- Carsim **Restreint au domaine des assurances pour les accidents de la route**
- Graph conceptuel à partir des besoins par Déclés

# Documents corporatifs

- Glossaire ou lexique pour les termes d'usage courant
- Définition de tâches, description de poste
- Code d'éthique
- Règlements internes, politiques, directives
- Procédures de traitements internes et externes
- Entente de relation avec les fournisseurs ou les clients
- Règles d'affaires
- Mission de l'entreprise, plan d'affaires, plan stratégique
- Tout autres documents décrivant des processus internes

# [ Type de bruit ]

- Éléments qui ne sont pas des concepts
  - Expressions vides composées de noms
    - « ... en matière de... »
    - « ...en l'occurrence... »
- Concepts indésirables (pour ce domaine)
- Concepts synonymes désignant une même entité
  - étudiant, personne, individu

# [ Sources du bruit ]

Le bruit, dans ce contexte, peut être produit par :

- Erreurs des outils de traitement
  - Mauvaises annotations  
Ex: verbe identifié comme un nom
- Pseudo synonyme :
  - « l'identificateur unique » vs « cette information »
- Trop grande couverture des annotations par rapport à l'efficacité des techniques d'élimination de bruit

# [ Outils utilisés dans le prototype ]

- Gate 3.1
  - Université de Sheffield, groupe NLP
- TreeTagger
  - Catégorisation des parties du discours
  - Probabiliste, multilingue
  - Université Stuttgart en Allemagne
- Sources linguistiques
  - Dictionnaires tiré d'Unitex
  - Liste de locutions