

Apprentissage et normalisation : une perspective textuelle

Laurent Romary

Laboratoire Loria-INRIA

Henri Hudrisier

Université de Paris VIII

Deux mondes séparés

- ◆ Un univers technologique en pleine évolution
 - Déploiement technologique et normatif
 - De nombreuses archives textuelles se mettent en place (OTA, WWP, Frantext, etc.)
- ◆ Des pratiques de recherche souvent inchangées dans les sciences humaines
 - Textes de références sous forme papier, difficiles d'accès
 - Des environnements d'échange inexistant (où est mon directeur de thèse?)
 - Un usage réduit des technologies (Word etc.)

Et pourtant:

- ◆ La recherche en sciences humaines pourrait être un lieu idéal d'utilisation des nouvelles technologies :
 - Accès nécessaire à de multiples sources d'information:
 - œuvres « classiques », dictionnaires de référence, productions scientifiques
 - La confrontation d'idées comme mode de production:
 - Rôle central de la glose et de l'argumentation

Rêvons un peu...

- ◆ Objectif:
 - Définir ce que serait un environnement intégré de recherche en sciences humaines
 - Application à la philosophie (Paris VIII)
 - Explorer ce que peuvent apporter les standards les plus récents
 - Evaluer la faisabilité d'un tel projet

Scénario

- ◆ Environnement de recherche de troisième cycle
 - Utilisateurs: étudiants et enseignants
 - Objectifs:
 - Appréhender des notions de base
 - Appréhender un univers documentaire
 - Accompagner la production scientifique

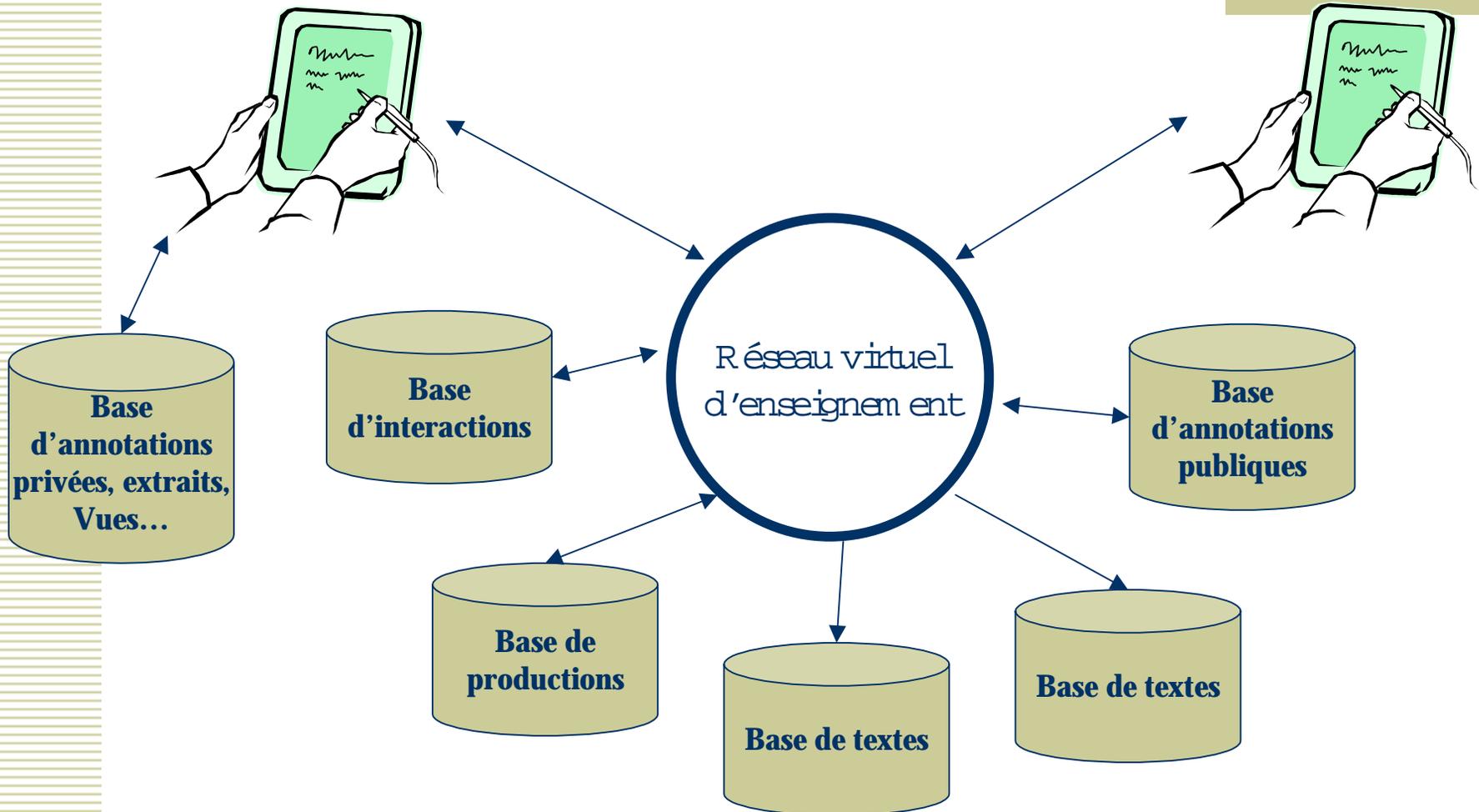
Fonctionnalités attendues

- ◆ Identification des ressources disponibles
 - Sources, courants/thèmes, degré de validation...
- ◆ Accès aux ressources
 - Lecture en ligne, accès ciblé à des concepts particuliers
- ◆ Edition et publications d'annotations
 - Notes personnelles, forum d'échange (tuteur, autres chercheurs)
- ◆ Croisement de sources multiples
 - Intertextualité (autorale, temporelle, thématique, linguistique, etc.), bases terminologiques (multilingues)

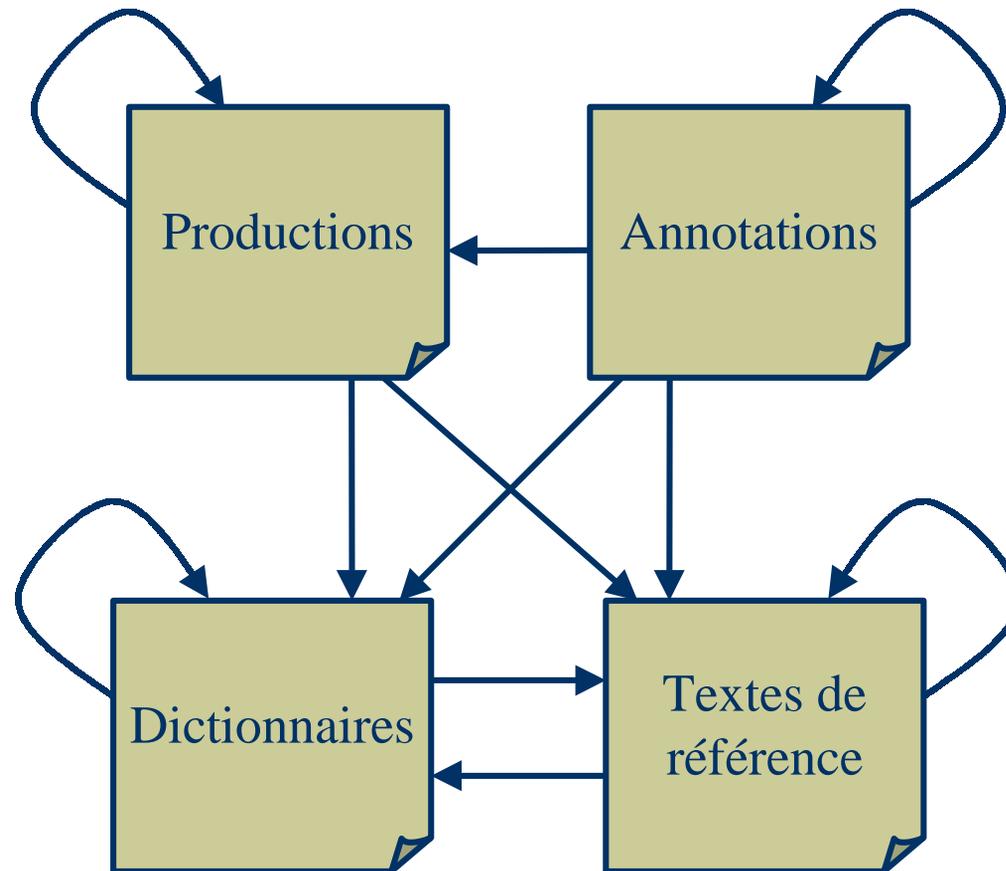
Besoins

- ◆ Textes structurés
 - Méta-données (auteur, titre, édition, concepts abordés/thématiques)
 - Organisation structurelle (division, paragraphes)
 - Annotation de surface (concepts, personnes, lieux, dates, bibliographie)
- ◆ Traitement homogène des textes classiques, dictionnaires et productions de recherche

Architecture



Un univers de liens



Un univers normalisé

- ◆ Pourquoi?
 - Garantir l'interopérabilité des différentes sources d'information
 - Cohérence des liens, transparence vis-à-vis de la source (identification commune des documents, poste de lecture unifié)
 - Disposer d'outils faciles à déployer
 - La philosophie ne disposera jamais du budget de l'industrie automobile...
 - Focaliser l'attention sur le contenu
 - Concepts, qualités des textes, scénarios d'interaction

Des normes ?

- ◆ Infrastructure (syntaxe de base, mécanismes de liens, services webs)
 - XML, RDF (annotations), SOAP/Services Web
- ◆ Textes, dictionnaires etc.
 - La TEI ! (Text Encoding Initiative)
- ◆ Terminologies multilingues
 - Normes du TC37/SC1 (méthodes) et TC37/SC3 (formats, ex. ISO 16642: TMF)
- ◆ Méta-données
 - TC37/SC4 (travaux issus de OLAC, ISLE)

Des modèles de document

- ◆ Derrière chaque format, il y a un modèle:
 - TEI : attachement d'un entête à chaque texte
 - ISO 16642/TMF : modèle conceptuel Wüstérien (identification des termes qui expriment un concept)
- ◆ Un espace à explorer (TC37/SC4): définir des modèles de schémas d'annotation:
 - Identification des liens référentiels (trace d'une personne, d'un concept): « l'auteur », « celui-ci », « il »
 - Actes de langage/modèles rhétoriques

Encore des normes?

- ◆ Apports possibles au JTC1/SC36
 - Expérimentation de scénarios d'interaction, de méta-données de qualification de sources documentaires (degré de validation, de pertinence)
- ◆ Apports possibles du JTC1/SC36
 - Etude de la place des technologies des langues dans un environnement pédagogique

Evaluation

- ◆ Quelques avantages :
 - Augmenter la « productivité » des thèses
 - Faciliter l'enseignement à distance
 - Accroître l'implication des universitaires dans la démarche normative
- ◆ Quelques difficultés :
 - Interfaces adaptées
 - Gestion de la qualité éditoriale (validation des productions, réduction de la redondance)
 - Droits associés aux textes/productions

Un rêve possible...

- ◆ L'environnement technique rend le déploiement d'un tel réseau possible
 - Matériel: Haut-débit, réseau sans fil, tablettes graphiques
 - Logiciel: XML devient de plus en plus transparent (serveurs, éditeurs, Annotea...)
 - Un dispositif ouvert rend un développement « en phases » accessible humainement et financièrement
- ◆ Il ne reste plus qu'à s'y mettre...