



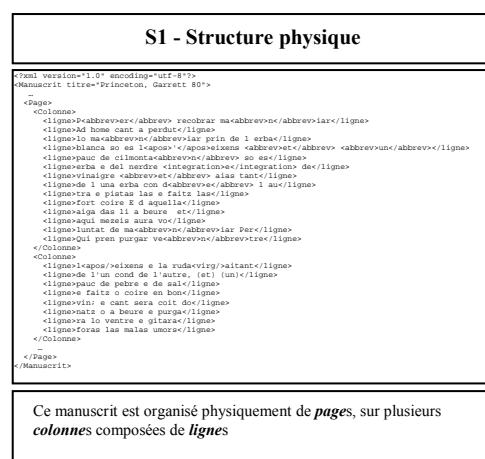
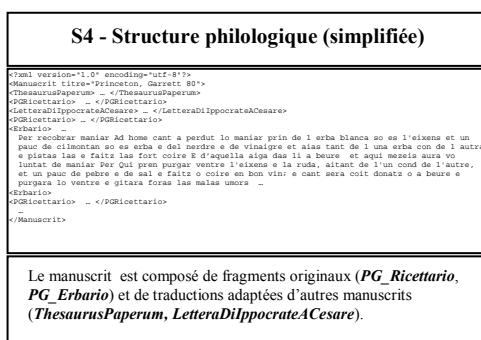
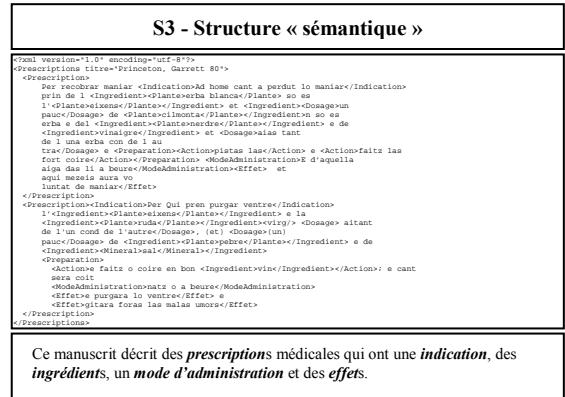
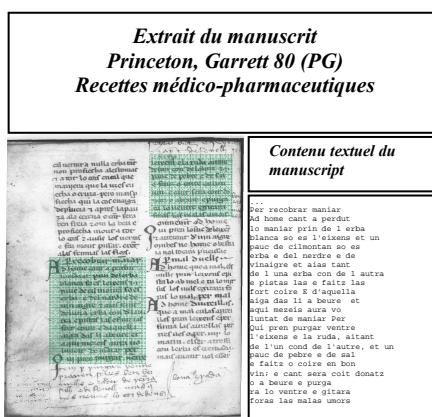
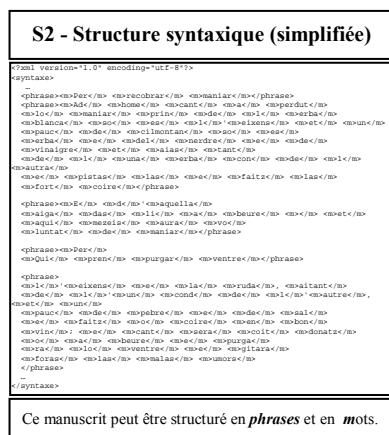
Documents multistructurés

*Représentation et manipulation dans
l'environnement XML*

Problématique générale

- Un même texte peut être étudié selon divers points de vue
 - Comment représenter ce type de documents possédant **plusieurs structures** ?
 - sur des fragmentations différentes du texte
 - sans relations hiérarchiques particulières (chevauchements possibles)
 - Comment manipuler conjointement l'ensemble de ces structures ?

Illustration sur des manuscrits médiévaux en occitan



- Quelle est la prescription qui comporte le plus de lignes ?
 - Quels sont les mots qui sont coupés en fin de ligne ?
 - Quelles sont les pages où apparaît entièrement une prescription provenant du manuscrit PG_Erbario ?
 - Combien y-a-t-il en moyenne de prescriptions par colonne ?

Objectifs

- Définir un ou plusieurs modèles formels :
 - Comparer des modèles à base d'arbres et de graphes
 - Etudier les syntaxes XML possibles
 - Proposer des opérateurs spécifiques et une extension de XQuery
 - Manipulation classique (sur structure et contenu) des instances étendue à la multistructure
 - Traitement des structures (comparaison/transformation)
 - Etudier le déploiement de documents multistructurés sur le Web

Remerciements

Dr M.-S. CORRADINI BOZZI, Philologue, Dipartimento di lingue e letterature romanze, Università di Pisa (Italia).

