

1 La génération du <body> grâce au lexique *SegmentOnto*

Expliquer comment j'ai généré le <body> en prenant les lignes de texte de la <sourceDoc> selon leurs étiquettes.

2 L'analyse linguistique

Parler d'un travail futur qui pourra prendre le fichier XML-TEI que j'ai créé et extraire les lignes de texte du <body> et les passer aux modèles TAL pour faire des analyses linguistiques. Montrer un exemple de ces données et comment cela marcherait avec un travail préliminaire que j'aurai fait d'ici la fin de stage et/ou le travail d'un ancien stagiaire que j'ai appliquée aux données Gallic(orpor)a.