

Tableau Pie (tagueur) + Pyrrha (interface de correction et de préannotation)

Accessibilité
- Téléchargement + Installation pour Pie ou Pyrrha
- Interface web pour Pyrrha avec possibilité d'exploiter un webservice Pie pour prétaguer : http://dev.chartes.psl.eu/ppa/
Interface
- Graphique pour Pyrrha (hors installation sur ordinateur personnel si souhaité)
- CLI (Command Line Interface) pour Pie ou API Web
Systèmes
- Linux et Mac pour Pie/Pyrrha en installation locale
- Tous pour Pyrrha installé sur un serveur web
Langage d'écriture du programme
Python
Langue(s)
- Disponible en ancien français, français classique, latin fin décembre, dépendant des jeux d'entraînement et des fichiers de configuration.
Jeux d'étiquettes
Dépend des jeux d'entraînement / Pour l'AF : Tobler-Lommatzsch – POS +Morph = Cattex 2009, pour Latin : LASLA Cf. https://github.com/hipster-philology/pyrrha/tree/dev/app/configurations/langs
Corpus d'entraînement
<p>Ancien Français Geste 100 000 mots ; J.B. Camps, Lucence Ing, Alice Cochet, Elena Albarran (et des M2) ; Corpus Juris Civilis env. 100 000 mots ; Frédéric Duval, Lucence Ing ; Chrestien 200 000 mots, fournis par Pierre Kunstmann, et alignés avec nos référentiels (JBC, LI) Wauchier 30 000mots ; Ariane Pinche ; Lancelot 15 000 mots ; Lucence Ing.</p> <p>Occitan Montferrand env. 30k mots ; Comptes des consuls de Montferrand, éd. Lodge ; JBC, Gilles Couffignal (Paris-Sorbonne), Marine Mazars ; Flamenca env. 20k mots, éd. P. Meyer, fournis par Olga Scrivner.</p> <p>Latin Cf. http://web.philo.ulg.ac.be/lasla/textes-latins-traites/</p>
Réentraînement/personnalisation (ou pas)
- Possible
Licence
- MIT pour Pyrrha
- Inconnue pour pie
Export/import – format(s)
- Plein texte vers TSV pour Pie
- Plein Texte ou TSV vers TSV (et bientôt TEI) pour Pyrrha

Dernière version/date
- Pie comme Pyrrha : Décembre 2018 (encore en développement actif, nouvelles fonctionnalités très souvent)
Site web
- https://github.com/hipster-philology/pyrrha
- https://github.com/emanjavacas/pie
Documentation (manuel)
- Pie : Non
- Pyrrha : Pas pour l'instant
Contacts/responsables
- Pyrrha : Thibault Clérice , Julien Pilla
- Pie : Enrique Manjavacas, Mike Kestemont
- Modèles préentraînés : JB Camps ou Thibault Clérice

On obtient jusqu'à 90% de réussite sur les textes en ancien français avec le modèle lemmes, POS et morphosyntaxe, voir les résultats ci-dessous :

Français moderne littéraire

- POS : 0.9769
- Lemma : 0.9499

Ancien Français (au 31 octobre) avec Pie

Corpus : Wauchier + Geste (~130.000 mots) uniquement

- POS: 0.9335
- Lemma : 0.9003
- Morph : 0.8468

Note : Un problème de configuration important sur les lemma doit être corrigé. Les résultats devraient augmenter suite à cette correction.