

Activités du Lundi 13 avril au Vendredi 17 mai :

- Transformation du dataframe id au format JSON (pour unifier avec le corpus en JSON)
- Récupération des events annotés dans les fichiers des dépendances (où les events n'étaient pas annotés)
- En attente : Récupération du target-path des events. (Réunion avec Anaïs et Gaëtan lundi 20 mai pour en discuter)
- Ajout des fonctions de création d'identifiants artificiels et de fusion des CSV dans le programme principal extractFeatures.py (car ils étaient dans deux programmes à part et, dans l'idéal, j'aimerais établir un unique script Python qui exécute tout.)
- Rédaction du rapport de stage :
 - Partie contexte :
 - Processus de tokenization
 - Partie des traits syntaxiques :
 - Définition de ce qu'est l'analyse en dépendance et constituance
 - Utilisation du Stanford Parser et d'addDiscourse (qui prend en entrée un arbre syntaxique)
 - Utilisation du Dependency Parser (à compléter quand la tâche de récupération du target-path sera finie)