Activités du Lundi 8 avril au Vendredi 12 avril :

- Formation linux le lundi 8 avril
    - Qu'est-ce qu'un OS ? comment ça fonctionne ?
    - Apprentissage des commandes linux de base (cd, ls, sudo, pwd, mkdir, rmdir, cp –R, mv, rm, cat, less…)
    - Petits exercices de manipulations de fichiers / recherche dans des fichiers

- Wordnet
    - Récupération des synsets pour les events avec nltk (qui permet d'utiliser wordnet sur python)
    - Récupération de l'hypéronyme de chaque event
    - Gaëtan suggérait d'utiliser les troponymes, Wordnet ne permet pas de les obtenir (à voir comment faire pour les avoir)
    - Est-ce que les antonymes seraient pertinents à utiliser ? (il est possible de les obtenir grâce à Wordnet)
    - Ajout dans le csv 'features_events.csv' des synsets / hypéronymes

- Nouveaux contextes
    - Ajout du contexte -4/+4 en POS
    - Ajout du contexte en stem (les stems sont peut-être suffisants pour analyser la ressemblance des termes dans le contexte, plutôt que d'utiliser des lemmes, ici on va juste regarder la racine plutôt que la forme canonique)

- Découverte de problèmes dans les timexs (fichier 'timex_contexts_id.csv')
    - Gaëtan a remarqué que des timex apparaissaient plusieurs fois
        - Le problème venait du fait que dans un timex il est possible d'avoir une virgule : March 22, 1999
        - Déjà dans le csv nous avions : March#t1>22 ,#t1>1999#t1
            - Alors que nous voulions : March#t1>22#t1>,#t1>1999#t1
                - Problème corrigé dans le programme 'tranformation_tml_to_txt.py'
        - Ensuite il y avait un problème dans le contexte de ces timex
            - Problème de tokenization à cause de la virgule, nltk tokenizer tokenize sur les virgules donc plusieurs contextes pour un seul timex (un pour March#t1>22#t1 et un pour ,#t1>1999#t1)
            - J'ai pris la décision de retirer les virgules dans ces timex car ils n'apportent pas d'informations supplémentaires. C'est juste de cette manière qu'est construite une date en anglais/américain…
                - Problème corrigé

- Remarque d'autres problèmes à résoudre