

## Activités du Lundi 4 janvier 2019 au Vendredi 8 Janvier 2019 :

### Réunion mardi 5 février 2019 :

- Discussion sur les features.
- Définition de la suite des événements : extraire les premiers features à partir des fichiers .tml (ce qui nous est proposé dans l'annotation).

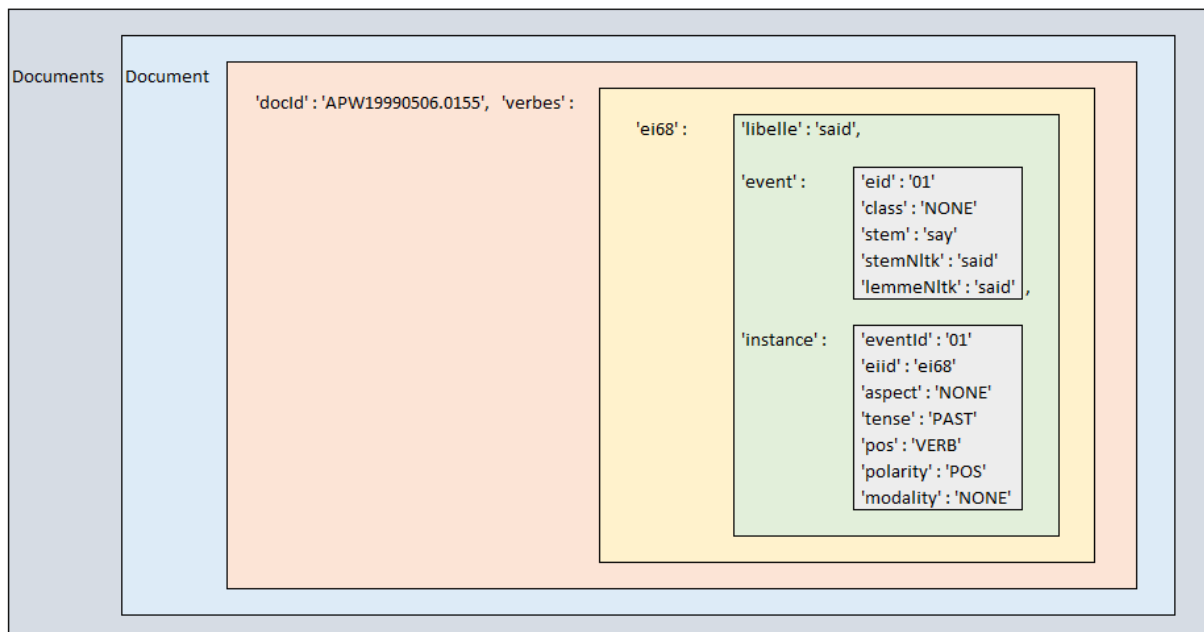
### Sur Python :

- Travail sur les fichiers du dossier TBAQ-cleaned > TimeBank
- **Extraction pour chaque event :**
  - o En début de fichier :
    - docID
  - o Event + ses attributs :
    - libelle de l'événement
    - eid
    - class
    - stem (pas présent dans tous les documents)
  - o Attributs de la balise MAKEINSTANCE
    - eiid
    - aspect
    - tense
    - pos
    - polarity
    - modality
  - o Ajout du stem et du lemme à l'aide du module NLTK
- **Création d'un fichier .CSV pour mettre en forme les données**

Pour chaque document nous avons les events + les instances + leurs attributs. La liste 'documents' englobe tous les documents. Le format est le suivant :

Documents [{document : {'docId' : 'wsj\_1038', verbes : {'ei1998' : {'libelle' : 'said', {'event' : {'eid' : '02', 'class' : ['OCCURRENCE'], 'stem' : 'say', 'stemNltk' : 'said', 'lemmeNltk' : 'said'}}}, {'instance' : {'eiid' : 'ei1998', 'aspect' : 'PERFECTIVE', 'tense' : 'PAST', 'pos' : 'VERB', 'polarity' : 'POS'}}}}}, {, document 2, etc... }]

En schéma :



### Event avec la classe « aspectual », quelques occurrences :

- Continue = continuer,
- begin = commencer,
- end = prendre fin,
- remain = rester,
- settlement = règlement,
- staunch = arrêter,
- clamp = serrer,
- stop = arrêter,
- finish = finir,
- renew = renouveler,
- return = revenir,
- keep = garder,
- become = devenir,
- unfolding = se dérouler,
- extending = extension,
- start = commencer,
- getting = obtenir.

Exemple :

« Saddam <EVENT eid="e30" class="REPORTING">said</EVENT> he will <EVENT eid="e31" class='aspectual'>begin</EVENT> withdrawing troops from Iranian territory »

- ➔ L'événement avec une classe dite aspectuelle donne une information sur la sémantique de l'événement.

### En cours :

- Visualiser la proportion des termes trouvés dans VerbeOcean.
  - o Comparaison de termes : Termes Event → Termes VerboOcean (relation happens-before)
  - o Quels termes sont en commun ? (356 termes en commun) / quels termes ne seront pas reconnus dans Verbocean ? (pour envisager d'utiliser des synonymes)
- Idée : lemmatiser les deux parties (ou stem), pour que les mots aient la même forme et puissent être comparés.

### En prévision :

- Effectuer un programme similaire pour les TIMEX (suggestion de Gaëtan).