**LES ESSENTIELS**

🡺 Sujet : « **Détection d’erreur dans un corpus anglais »**

l’idée est d’annoter un texte en anglais pour y indiquer les mots comportant des fautes d’orthographes, ainsi que les fautes de grammaire (qui peuvent être par exemple l’emploi d’une préposition erronée). Le corpus à utiliser est [ilexir](https://www.ilexir.co.uk/datasets/index.html) . Dans la mesure où le texte peut comporter des fautes d’orthographes, on pourra explorer en particulier le problème de la représentation des mots (par embedding, comme suite de caractères, ou avec un système mixte), et voir les influences sur la qualité du résultat.

* Le projet consistera **à traiter un problème de modélisation** (avec un jeu de données) , à mettre au point ou analyser des modèles prédictifs s’inscrivant dans la continuité des méthodes décrites dans les blocs de l’UE RCP217. Il convient dans le projet de **justifier** les choix de méthode, de paramètres, etc. ; lier les résultats des éventuelles analyses préalables au choix de la ou des méthodes de modélisation.

Le rapport devra inclure une présentation claire mais concise du problème et des données, une définition de l’objectif de l’étude, la démarche choisie, sa formalisation et les résultats obtenus, ainsi qu’une conclusion, **sans dépasser les 20 pages**. D’éventuelles illustrations complémentaires, non essentielles pour la compréhension du travail, pourront être mises dans des annexes (après les ≤20≤20 pages du rapport).

* Le rapport **doit être accompagné des programmes que vous avez développés** **devra être archivé dans un seul fichier .zip** . Attention, la taille maximale du fichier est de 20 Mo.

**Table des matières RCP217**

* [Présentation](https://cedric.cnam.fr/vertigo/Cours/RCP217/preambule.html)
  + [Public et conditions d’accès](https://cedric.cnam.fr/vertigo/Cours/RCP217/preambule.html#public-et-conditions-d-acces)
  + [Objectifs pédagogiques](https://cedric.cnam.fr/vertigo/Cours/RCP217/preambule.html#objectifs-pedagogiques)
  + [Compétences visées](https://cedric.cnam.fr/vertigo/Cours/RCP217/preambule.html#competences-visees)
  + [Contenu](https://cedric.cnam.fr/vertigo/Cours/RCP217/preambule.html#contenu)
  + [Modalité d’évaluation](https://cedric.cnam.fr/vertigo/Cours/RCP217/preambule.html#modalite-d-evaluation)
  + [Équipe enseignante](https://cedric.cnam.fr/vertigo/Cours/RCP217/preambule.html#equipe-enseignante)
* [Données spatio-temporelles (6 séances)](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursDonneesSpatiotemporelles.html)
  + [Apprentissage incrémental, données en flot](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursDonneesSpatiotemporelles.html#apprentissage-incremental-donnees-en-flot)
  + [Données séquentielles, réseaux RNN, données vidéo](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursDonneesSpatiotemporelles.html#donnees-sequentielles-reseaux-rnn-donnees-video)
  + [Apprentissage profond pour les données audio](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursDonneesSpatiotemporelles.html#apprentissage-profond-pour-les-donnees-audio)
  + [Apprentissage profond et séries temporelles](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursDonneesSpatiotemporelles.html#apprentissage-profond-et-series-temporelles)
* [Traitement de la langue naturelle](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursTraitementLangueNaturelle.html)
  + [Cours](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursTraitementLangueNaturelle.html#cours)
* [Systèmes de recommandation, apprentissage sur graphes (GraphML)](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursRecommandationGraphML.html)
  + [Introduction aux systèmes de recommandation](https://cedric.cnam.fr/vertigo/Cours/RCP217/coursRecommandationGraphML.html#introduction-aux-systemes-de-recommandation)
* [Page de recherche](https://cedric.cnam.fr/vertigo/Cours/RCP217/search.html)
* [Index](https://cedric.cnam.fr/vertigo/Cours/RCP217/genindex.html)