
N° 2 | 2024
2024

Insertion des catégories grammaticales dans le Dictionnaire Électronique des Synonymes (DÉS)

Laurette CHARDON *Ingénieure de recherche*
UFR HSS
CRISCO
Université de Caen Normandie

Édition électronique :

URL :

<https://demc-journal.org/articles/revue-2/2790-insertion-des-categories-grammaticales-dans-le-dictionnaire-electronique-des-synonymes-des>

DOI : numerev_2191

ISSN : 3036-5295

Date de publication : 31/12/2024

CertiScience® *Certifié évalué par les pairs*

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : CHARDON, L. (2024) Insertion des catégories grammaticales dans le Dictionnaire Électronique des Synonymes (DÉS). *DEMC Journal*, (2).

https://doi.org/10.34745/numerev_2191

Cet article détaille la procédure suivie pour introduire la catégorie grammaticale dans la base de données du [Dictionnaire Électronique des Synonymes](#) (DÉS) du laboratoire [CRISCO](#) (50 457 entrées au 30 janvier 2024). Ce travail a eu lieu en deux étapes : une première de janvier à novembre 2022 et une seconde de juin à novembre 2023. La première étape de janvier à novembre 2022 s'est déroulée en deux grandes phases. A partir d'un premier jeu de données de l'[ATILF](#) (Analyse et Traitement Informatique de la Langue Française, UMR7118), laboratoire du CNRS chargé de la maintenance et du développement du TLFi (Trésor de la Langue Française informatisé), transmis sous forme d'un classeur (au format .xlsx) contenant pour chaque entrée, les catégories grammaticales correspondantes. Avec des programmes en langage Python, les verbes, substantifs, adjectifs et adverbes ont été introduits dans la base de données du DÉS. Ensuite une recherche par schémas (-ais, -euse, -ale, -ande, . . .) avec plusieurs groupes de mots mélangés. A partir d'autres moyens : le modèle d'intelligence artificielle (IA) `fr_dep_news_trf` disponible avec la librairie Spacy en langage Python va permettre au moyen de l'étiquetage morpho-syntaxique d'associer aux entrées du DÉS des informations grammaticales. La seconde étape de juin à novembre 2023 s'est déroulée en deux grandes phases : l'étude d'un second et ancien jeu de données de l'ATILF contenant 49 855 entrées avec les catégories grammaticales ; l'étude d'un fichier plus récent (septembre 2023) de l'ATILF de 103 329 lignes. Cet article s'appuie sur un document de travail, accompagné d'un dépôt git public.

This article details the procedure followed to introduce the grammatical category into the CRISCO laboratory's [Dictionnaire Électronique des Synonymes](#) (DÉS) database (50,457 entries, January 30th 2024). This work was carried out in two stages: the first from January to November 2022 and the second from June to November 2023. The first stage, from January to November 2022, took place in two main phases. First using an initial dataset from ATILF (Analyse et Traitement Informatique de la Langue Française, UMR7118), the CNRS laboratory responsible for maintaining and developing the TLFi (Trésor de la Langue Française informatisé), transmitted in the form of a workbook (.xlsx format) containing the corresponding grammatical categories for each entry. Using Python language programs, verbs, nouns, adjectives and adverbs were entered into the DÉS database. This was followed by a search using patterns (-ais, -euse, -ale, -ande,) with several groups of words mixed together. Then by other means: the artificial intelligence (AI) model `fr_dep_news_trf` available with the Spacy library in the Python language will use morpho-syntactic labelling to associate grammatical information with DES entries. The second stage, from June to November 2023, took place in two main phases: the study of a second, older ATILF dataset containing 49,855 entries with grammatical categories ; the study of a more recent ATILF file (September 2023) containing 103,329 lines. This article is based on a working document, accompanied by a public git repository.

