

Qu'est-ce que le DES ?

La maintenance du DES
DES et la science ouverte

1er focus : graphe d'adjacence, cliques et espace sémantique

2nd focus : l'insertion des catégories grammaticales

Références

Présentation du Dictionnaire Électronique des Synonymes (DES)

Laurette CHARDON

Journée d'étude de l'axe 2 du laboratoire CRISCO - UR 4255 - Université de Caen

8 juin 2023

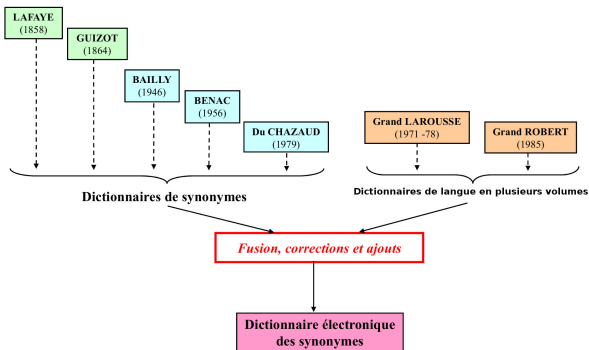


1/31

Historique^[1] I

- **Conception 1996-1999** : Bernard Victorri, mathématicien et linguiste CNRS et Sabine Ploux, ingénieure CNRS
- **Principe de base** : Représenter un dictionnaire aussi complet que possible sous forme de **graphe** dont les **sommets** sont les entrées et les **arêtes** les liens synonymiques pour créer un espace sémantique en traitant mathématiquement divers aspects de la polysémie d'une vedette.
- **La base de données** : issue de l'INALF (Institut national de la langue française) constituée de 7 dictionnaires avec un travail d'homogénéisation et de symétrisation.

Historique^[1] II



Historique^[1] III

- La **symétrisation** signifie que si l'entrée A est synonyme de l'entrée B alors l'entrée B est synonyme de l'entrée A. La synonymie est définie ainsi : s'il existe **au moins un contexte** dans lequel on peut substituer B à A sans en modifier significativement le sens, alors il en existe au moins un dans lequel on peut substituer A à B.
- Le fondement des calculs sur le graphe pour l'étude de la polysémie est la notion de **clique**.

Qu'est-ce que le DES ?

La maintenance du DES
DES et la science ouverte

1er focus : graphe d'adjacence, cliques et espace sémantique
2nd focus : l'insertion des catégories grammaticales

Références

L'historique

L'interface actuelle et celle de démonstration

les interfaces I

<http://crisco.unicaen.fr/des>



Actual DES - Accès direct DES - Kall

Dictionnaire Electronique des Synonymes (DES)

Dernières Actualités **la lettre d'actualité n°6 du DES vient de paraître !**

Tapez l'unité lexicale recherchée puis cliquez sur Valider ou tapez sur Entrée (données à jour du 26 août 2019)

Rechercher : Valider

curieux définition  espace sémantique 

55 synonymes

étonnant, étrange, amateur-trice, amusant, anxieux, attachant, atterré, évidé, badaud, beau, bizarre, bizarroïde, chercheur, collectionneur-euse, déconcertant, désireux, dilettante, drolatique, drôle, extraordinaire, extravagant, fâcheux, fouille-au-pot, fouilleux, fouinard, fouineux, fureteur, inattendu, incompréhensible, inconcevable, incroyable, indiscret, inouï, inquisiteur, insolite, intéressé, intéressant, intrigué, inusité, investigateur, juge, original, paradoxal, piquant, pittoresque, plaisant, rare, regardant, singulier, sourcilieux, spectateur-trice, surprenant, tuculent, unique, voyeur

18 antonymes

étourdi, acrotrope, anodin, banal, blasé, commun, courant, discret, habituel, incurieux, indifférent, insouciant, ordinaire, quelconque, quotidien, réservé, usuel, vulgaire



5/31

les interfaces II

Classement des premiers synonymes

étrange	■
bizarre	■
original	■
drôle	■
fureteur	■
fouineur	■
déconcertant	■
inquisiteur	■
pittoresque	■
singulier	■
truculent	■
insolite	■
paradoxal	■
étonnant	■
amusant	■
chercheur	■

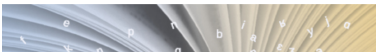
59 cliques

- étonnant, étrange, bizarre, curieux, drôle, extraordinaire, insolite, original, singulier, surprenant
- étonnant, étrange, bizarre, curieux, extraordinaire, insolite, inusité, original, rare, singulier
- étonnant, étrange, bizarre, curieux, drôle, extraordinaire, original, singulier, truculent
- étonnant, étrange, bizarre, curieux, extraordinaire, original, paradoxal, singulier, surprenant
- étonnant, étrange, curieux, extraordinaire, inconcevable, incroyable, inouï, paradoxal, surprenant
- étrange, bizarre, curieux, drôle, extraordinaire, extravagant, insolite, original, singulier
- étonnant, étrange, bizarre, curieux, extraordinaire, incroyable, paradoxal, surprenant
- bizarre, curieux, drôle, extraordinaire, insolite, original, pittoresque
- bizarre, curieux, drôle, extraordinaire, original, pittoresque, truculent
- bizarre, curieux, extraordinaire, extravagant, original, singulier, unique
- bizarre, curieux, extraordinaire, original, rare, singulier, unique
- étonnant, étrange, bizarre, curieux, déconcertant, inusité, original

les interfaces III

<http://crisco3.unicaen.fr/des>

DÉS · DICTIONNAIRE ÉLECTRONIQUE DES SYNONYMES CRISCO UNICAEN



DÉCOUVREZ

Le dictionnaire électronique des synonymes du CRISCO (DÉS) contient aujourd'hui plus de 50 000 entrées et 209 000 relations synonymiques réciproques. La base de départ a été constituée à partir de sept dictionnaires classiques. Un premier travail, réalisé par l'INALF (Institut National de la Langue Française), a permis d'en extraire les relations synonymiques. Le laboratoire ELSAP, qui est devenu par la suite le CRISCO, a ensuite concaténé, homogénéisé et symétrisé les données. Depuis 1994, un important travail de correction se perpétue par l'ajout ou la suppression de liens synonymiques et antonymiques.

[En savoir plus](#)



les interfaces IV

DÉS · DICTIONNAIRE ÉLECTRONIQUE DES SYNONYMES

CRISCO UNICAEN

× curieux



CURIEUX



regroupement ▾

55 SYNONYMES



ORDRE : ● ↕ ○ ⇅

amateur-trice | amusant | anxieux | attachant | attentif
| avide | badaud | beau | bizarre | bizarroïde |
chercheur | collectionneur-euse | déconcertant |
désireux | dilettante | drolatique | drôle | étonnant |
étrange | extraordinaire | extravagant | flâneur |
fouille-au-pot | fouilleur | founard | fouineur |
fureteur | inattendu | incompréhensible | inconcevable
| incroyable | indiscret | inouï | inquisiteur | insolite |
intéressant | intéressé | intrigué | inusité |
investigateur | juge | original | paradoxal | piquant |
pittoresque | plaisant | rare | regardant | singulier |
soucieux | spectateur-trice | surprenant | truculent |
unique | voyeur

18 ANTONYMES

acratopège | anodin |
banal | blasé | commun
| courant | discret |
étourdi | habituel |
Incurieux | Indifférent |
insouciant | ordinaire |
quelconque | quotidien |
réservé | usuel | vulgaire

1 SONORITÉ PROCHE - KURI

curie



Les étapes mensuelles

- 1 Traitement et saisie des propositions des utilisateurs
- 2 Traitement et saisie des propositions calculées automatiquement par programme
- 3 Extraction de la base, traitements, vérifications, mise en ligne, sauvegardes
- 4 **Mise à jour des statistiques** (consultations, ajouts, suppressions, modifications, contributeurs)
- 5 **Traitement des mots les plus recherchés** (journalier)
- 6 Mise à jour du DES autonome

Méthodologie suivie pour tout ajout dans la base de données

- Recherche d'une phrase dans laquelle l'un des termes peut remplacer l'autre et inversement \Rightarrow symétrie et contexte
- Même type d'objet (animé ou inanimé), même catégorie grammaticale
- Recherche des définitions dans le CNRTL, le Robert (en ligne via le SCD), wiktionary, parfois Frantext
- Exemples retenus : *référence - sommité ; essai-proposition ; facteur-paramètre ; hypothèse - supposition*
- Exemples non retenus :
 - pas dans les dictionnaires : *défourer # déterrer, sérieux # agelaste*
 - pas de contexte : *récolte # enlèvement, engager # produire*
 - catégorie grammaticale différente : *frime # clinquant*

Statistiques

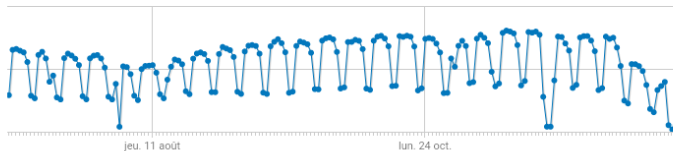
De la base de données :

- Le nombre d'entrées est \approx de 50.300
- Le nombre de liaisons synonymiques est \approx de 212.200
- Le nombre de cliques est \approx de 111.700
- l'entrée qui a le plus grand nombre de synonymes : **BON avec 242**
- l'entrée qui a le plus d'antonymes : **CALME avec 192**

D'accès (2022) :

- Le nombre de visites \approx 4.640.000, de pages vues \approx 40.540.000
- 167 pays différents (Europe, Amérique du Nord, Asie, ...)

Statistiques journalières d'accès

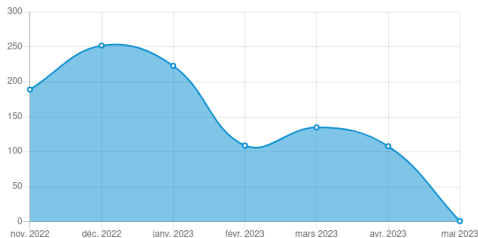


- les mots les plus recherchés,
- le classement des premiers synonymes,
- la féminisation,
- le calcul des liens probables,
- les liaisons litigieuses,
- la distance du khi-deux,
- le profil des utilisateurs, leurs remerciements, etc...

⇒ Voir le détail dans la présentation du DES ^[2] ([halshs-02489368](https://halshs.archives-ouvertes.fr/halshs-02489368))

Les données du DES sont sur la plate-forme **ORTOLANG**

Statistiques de consultation (1017 vues au total)



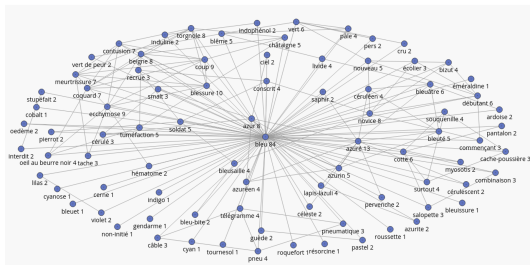
Statistiques de téléchargements (depuis novembre 2016 🇫🇷)

35 téléchargements complets
12 fichiers ou dossiers téléchargés

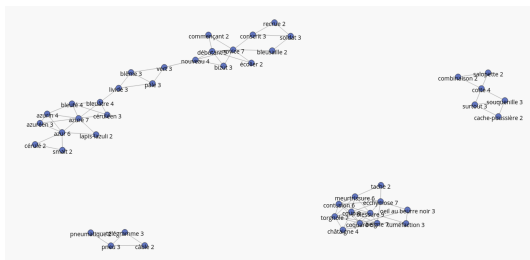
Le Plan de Gestion de Données : en cours de rédaction
Un **dépot git** avec les données des publications réalisées.

**Matrice
d'adja-
cence de
BLEU(84
synonymes) :**
représentation
graphique des
liens
synonymiques
enregistrés
dans la base
de données
sans aucun
calcul

Graphe de BLEU



Matrice d'adjacence de BLEU avec une présentation "améliorée" :
tous les liens avec la vedette ont été supprimés ainsi que les synonymes isolés



Interface : crisco3.unicaen.fr/desgraphe

- choix suppression de la vedette
- choix d'affichage des synonymes avec un nombre de liens supérieur à une limite choisie
- possibilité de sélectionner plusieurs mots vedette
- choix d'une méthode de regroupement (multilevel, infomap, spinglass)
- visualisation 3D

Graphes d'adjacence du DES - CRISCO - Université de Caen

Le graphe d'adjacence d'une vedette V donnée (ou d'un groupe de vedettes) est une représentation graphique de tous les liens synonymiques de V et également des liens entre les synonymes. Chaque sommet apparaît avec le mot qu'il représente suivi du nombre de liens que ce sommet possède avec les autres sommets.

Choisir un type d'affichage :

3D 2D Clusters 3D 3D Clusters

Fichiers à utiliser

Fichiers Mars 2022

Vedettes à saisir séparées par " " (sans espace):

liaisoncoeur

Générer aléatoirement

n_clicks : "0"

Essayer l'affichage de la(s) vedette(s)

Aide

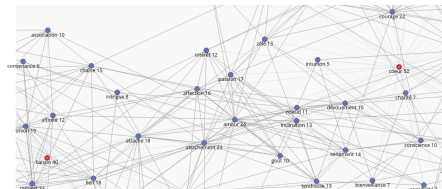
Essayer l'affichage des arêtes de la(s) vedette(s)

Suppression des sommets qui ont moins de N liens :

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Aide

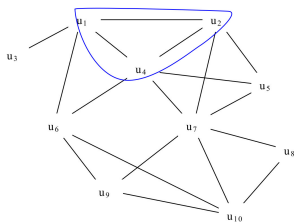
Graphe de LIAISON;COEUR (141 synonymes au total - 91 sommets affichés)



Calcul des cliques^[3]

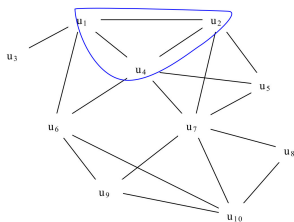
- Une clique (sous-graphe complet maximal) est un ensemble le plus grand possible de sommets du graphe tous reliés deux à deux, « le plus grand possible »
- En d'autres termes, une clique est un ensemble d'unités lexicales (= sommet) qui ont la propriété d'être toutes synonymes les unes des autres, aucune autre unité ne pouvant être rajoutée à l'ensemble sans que l'on perde cette propriété.

Exemple :



Le sous-ensemble $\{u1, u2, u4\}$ est-il une clique ?

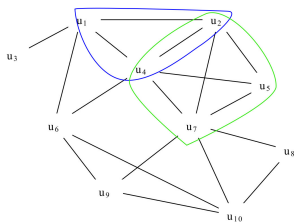
Exemple :



Le sous-ensemble $\{u_1, u_2, u_4\}$ est-il une clique ?

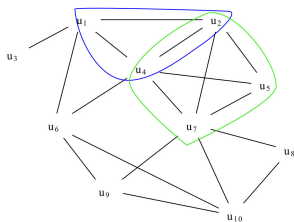
⇒ **OUI!**

Exemple :



Le sous-ensemble $\{u2, u4, u5, u7\}$ est-il une clique ?

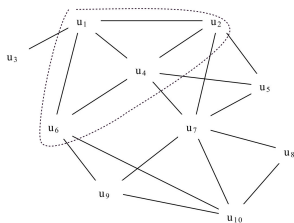
Exemple :



Le sous-ensemble $\{u_2, u_4, u_5, u_7\}$ est-il une clique ?

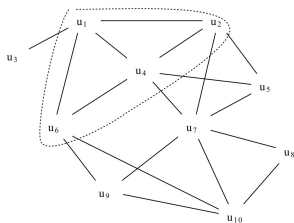
⇒ **OUI!**

Exemple :



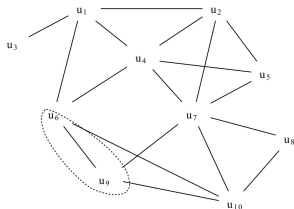
Le sous-ensemble $\{u1, u2, u4, u6\}$ est-il une clique ?

Exemple :



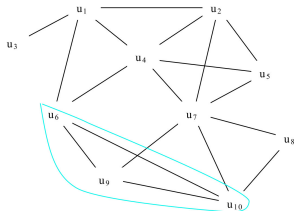
Le sous-ensemble $\{u1, u2, u4, u6\}$ est-il une clique ?
 \Rightarrow **NON !** puisque $u2$ et $u6$ ne sont pas synonymes.

Exemple :



Le sous-ensemble $\{u_6, u_9\}$ est-il une clique ?

Exemple :



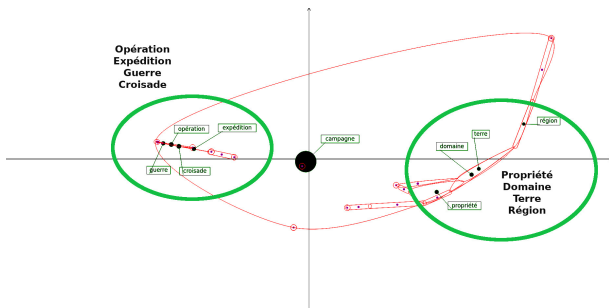
Le sous-ensemble $\{u6, u9\}$ est-il une clique ?

⇒ **NON !** parce qu'elle n'est pas maximale : ces deux unités sont toutes deux synonymes de $u10$ et c'est donc $\{u6, u9, u10\}$ qui constitue une clique.

- On réalise une Analyse Factorielle des Composantes (AFC)
- A partir de la matrice avec en ligne les cliques et en colonne les entrées
- Elle va permettre d'extraire l'essentiel de l'information sur un grand nombre de données (impossible à voir à l'œil nu) au travers de quelques facteurs les plus représentatifs.

Espace sémantique : exemple avec CAMPAGNE

Deux composantes se dégagent : “stratégique” (opération, expédition, guerre, croisade) et “topographique” (propriété, domaine, terre, région) :



Accès en ligne : <https://crisco.unicaen.fr/espsem/campagne>

La catégorie grammaticale I

- Le projet a duré 2 ans (janvier 2021 à novembre 2022)
- A partir des données de l'ATILF (*Analyse et Traitement Informatique de la Langue Française, UMR7118*)
- Les étapes :
 - ① Étude préliminaire du fichier de départ du TLFi (54.280 lignes),
 - ② Traitements informatiques avec le fichier TLFi,
 - ③ Traitements informatiques sans le fichier TLFi,
 - ④ Traitements manuels
 - ⑤ Vérifications

La catégorie grammaticale II

Col 1	Col 2	Col3	Col4	Col5
ABATTRE ABBATTRE ABATTU ... ABORAL ... ABOTÉ ... AUTO(-)DESTRUCTEUR ... AUTOPORTANT ... MÉTROPOLITAIN1 MÉTROPOLITAIN2 MÉTROPOLITE METS METTABLE METTEUR METTON METTRE MIS MISE MÉTURE MEUBLANT MEUBLE1 MEUBLE2 MEUBLE3	ABBATTRE voir ABATTRE UE ALE ÉE TRICE ANTE -AINE -AINE subst. masc. subst. masc. adj. -EUSE subst. masc. verbe part. passé et adj. subst. fém. -ANTE subst. masc. adj. adj. et subst.	ABATRE AUX ABOTTÉ AUTO(-) DESTRUCTIF AUTOPORTEUR adj. subst. masc. et adj. subst. part. prés. et adj.	verbe trans. adj. ÉE IVE EUSE MÉTRO	 adj. adj. adj. adj. et subst. subst. masc.

Table 1: Extrait du fichier de départ du TLFi

La catégorie grammaticale III

→ Gérer plusieurs types de lignes

- Étape 1 : traitements avec le fichier TLFi selon la catégorie grammaticale
 - les verbes (6.982 lignes)
 - les adjectifs (4.641 lignes)
 - les substantifs (28.588 lignes)
 - les adverbes (943 lignes)
 - plusieurs traitements sur des " mots mélangés " selon leur terminaison : -ante, -ine, etc..
- Étape 2 : traitements sans le fichier TLFi avec la librairie Spacy de Python sur les 16.000 entrées du DES non renseignées
 - 5 extractions du résultat suivant les codes retournés : VERB, NOUN, ADJ

La catégorie grammaticale IV

- Étape 3 : traitements manuels sur les 8.500 entrées restantes en utilisant des schémas (commençant par *de*, *sur*, *avec*, un verbe...) en 19 étapes.
 - Vérifications : par exemple, d'un coté un verbe (ni substantif, ni adjectif, ni adverbe, ni locution) et de l'autre un mot non verbe, ni adverbe, ni locution
 - *fier* : verbe mais aussi adjectif
 - *pourparler* : non verbe mais substantif masculin
 - *conseiller* : verbe mais aussi substantif
- 250 corrections réalisées

⇒ 244 codes grammaticaux (*adj.*, *adj. composé*, *adj. et adv.*, *adj. et interj.*, ...) et 430 combinaisons de codes grammaticaux.

⇒ Voir le rapport ^[4] [halshs-03956407](#) et le [dépôt git correspondant](#)

- [1] Michel MOREL et Jacques FRANÇOIS. “Le Dictionnaire Electronique des Synonymes du CRISCO : un outil de plus en plus interactif”. In : *Revue française de linguistique appliquée* Vol. XX.39 (2015), p. 9-28. URL : <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2015-1-page-9.htm#>.
- [2] CRISCO. *Présentation du DES : le bilan des mises à jour*. URL : <http://crisco.unicaen.fr/dictionnaire-electronique-des-synonymes/presentation-du-des/>.

- [3] Sabine PLOUX et Bernard VICTORRI. “Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes”. In : *Traitement Automatique des Langues* 39 (1998), p. 161-182. URL : <https://halshs.archives-ouvertes.fr/halshs-00009433>.
- [4] Laurette CHARDON. *Insertion des catégories grammaticales dans le Dictionnaire Électronique des Synonymes (DES) - Document de travail. Rapp. tech.* URL : <https://shs.hal.science/halshs-03956407>.