



**HAL**  
open science

# Éléments de recherche extraits des lettres d'actualités du Dictionnaire Électronique des Synonymes (DES) du CRISCO de janvier 2018 à février 2023

Laurette Chardon

## ► To cite this version:

Laurette Chardon. Éléments de recherche extraits des lettres d'actualités du Dictionnaire Électronique des Synonymes (DES) du CRISCO de janvier 2018 à février 2023. 2023. hal-04137377v2

**HAL Id: hal-04137377**

**<https://hal.science/hal-04137377v2>**

Preprint submitted on 19 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# CRISCO

Langues | Signification | Contexte

## Éléments de recherche extraits des lettres d'actualités du Dictionnaire Electronique des Synonymes (DES) du CRISCO

de la lettre n°1 (janvier 2018) à la lettre n°19 (Février 2023)

Laurette CHARDON - <https://cv.hal.science/laurette-chardon>

19 mars 2024

- Les principes composant la méthode de mise à jour du DES
  - Le principe de synonymie appliqué
  - La synonymie partielle versus pure
  - La gestion des participes passés
  - Les variantes féminines et masculines
  - Le traitement des synonymes isolés
  - Les relations litigieuses
  - Les expressions figées
- Les outils connexes
  - Le graphe d'adjacence
  - L'espace sémantique
- Les mots étudiés
  - mutin, facétieux, factieux, mdr et lol
  - Responsable
  - Curieux
  - Gagner
  - Travail
  - Les termes utilisés lors des voeux du nouvel an
  - Bien
  - Signe, Icône
  - Maison
  - Schéma, diagramme, graphique, graphe et carte
  - Nation, Culture
- Les mots les plus recherchés et les mots inexistantes
  - Les mots les plus recherchés - quelques statistiques
  - Les mots inexistantes
  - Les recherches particulières

Ce document reprend les éléments de recherche développés dans les lettres d'actualités n°1 à n°19 du DES : les principes sur lesquels reposent la mise à jour du DES, les outils connexes, les mots étudiés, les mots les plus recherchés

## Les principes composant la méthode de mise à jour du DES

### Le principe de synonymie appliqué

**Avril 2018 - Actudes2**

Il est opportun de rappeler le principe utilisé pour définir s'il existe ou pas un lien synonymique entre deux mots notamment ceux qui sont dans vos propositions. Comme cela est précisé sur la page Bref historique du DES, les initiateurs du projet sont Bernard Victorri et Sabine Ploux. Il ont rédigé plusieurs publications à ce sujet et en particulier celle intitulée [Construction d'espace sémantiques à l'aide de dictionnaires de synonymes](#) ou ils définissent entre autres les différents cas de synonymie à savoir la **synonymie pure** et la **synonymie partielle**.

C'est cette dernière qui est prise en compte dans le DES. La définition d'une synonymie partielle est la suivante : Deux unités lexicales sont en relation de synonymie si toute occurrence de l'une peut être remplacée par une occurrence de l'autre dans un certain nombre d'environnements sans modifier notablement le sens de l'énoncé dans lequel elle se trouve.

Cette définition, reprise dans cette autre publication [Quand les mots s'organisent en réseaux](#) est illustrée par un exemple :

*"Pour ne prendre qu'un exemple, défendre et interdire sont synonymes parce qu'ils sont à peu près équivalents dans des énoncés tels que défendre de fumer et interdire de fumer. De même, défendre et soutenir sont synonymes parce que défendre les droits de l'homme et soutenir les droits de l'homme ont sensiblement le même sens. En revanche, interdire et soutenir ne sont pas synonymes car il n'existe pas de contexte dans lequel on puisse les permuter sans modifier considérablement le sens de l'énoncé qui subit la transformation. On en déduit que défendre n'a pas le même sens dans tous les énoncés puisqu'il est remplaçable tantôt par interdire et tantôt par soutenir, qui n'ont, eux, jamais le même sens."*

### La synonymie partielle versus pure

**Octobre 2020 - Actudes10**

La synonymie que nous réalisons dans le DES est "partielle". Elle a été définie par Bernard Victorri et Sabine Ploux les concepteurs du DES de la façon suivante (issue de cette publication) :

*“Deux unités lexicales sont en relation de synonymie si toute occurrence de l’une peut être remplacée par une occurrence de l’autre dans un certain nombre d’environnements sans modifier notablement le sens de l’énoncé dans lequel elle se trouve.”*

Contrairement à la synonymie “pure” qui est définie par :

*“Deux unités lexicales sont en relation de synonymie pure si toute occurrence de l’une peut être remplacée par une occurrence de l’autre dans tout environnement sans modifier notablement le sens de l’énoncé dans lequel elle se trouve.”*

La synonymie “partielle” est beaucoup moins restrictive. En effet, il existe très peu de synonymes “purs” en raison de la polysémie largement répandue dans le langage. Une synonymie pure implique une propriété mathématique appelée transitivité : si A est synonyme de B et B synonyme de C, alors A est synonyme de C.

Dans le cas d’une synonymie partielle, et en raison de la polysémie, cette propriété n’est pas systématique vérifiée. Par exemple, *défendre* et *interdire* sont synonymes dans le contexte *défendre /interdire de fumer*. *Défendre* et *soutenir* le sont dans *défendre/soutenir les droits de l’homme* mais *soutenir* et *interdire* ne sont en aucun cas synonymes. Cela dit, il est important de se rappeler que même en cas de synonymie partielle, la catégorie grammaticale est à respecter : un adjectif “pur” ne peut pas être remplacé par un nom “pur” car il doit exister un contexte dans lequel l’un peut remplacer l’autre. Donc l’un des deux doit appartenir aux deux catégories grammaticales pour être éventuellement synonymes.

## La gestion des participes passés

### **Octobre 2020 - Actudes10**

L’ajout d’un participe passé issu directement d’un verbe n’est pas systématique. Par exemple, nous avons reçu comme propositions:

- *fixé* synonyme de *observé*
- *fixé* synonyme de *scruté*
- *fixé* synonyme de *épié*
- *fixé* synonyme de *averti*
- *fixé* synonyme de *concentré*
- *fixé* synonyme de *hypnotisé*

En effet, *fixé* est bien synonyme de tous ces mots mais aucun n’est un adjectif “pur”. Il n’y a pas d’intérêt à faire correspondre la forme participe passé d’un verbe à toutes les formes participe passé des synonymes de ce verbe : c’est redondant et n’apporte rien de plus puisqu’il suffit de rechercher les synonymes de la forme verbale pour obtenir tous les synonymes du type participe passé.

Par contre, *fixé* a été ajouté car il est synonyme de certain, normal, inséparable qui sont des formes uniquement adjectivales. Quand 2 formes

participe passé sont proposées comme synonymes et si elles sont déjà présentes dans la base de données, nous ajoutons le lien synonymique mais nous ne créons pas une nouvelle entrée de type participe passé si le seul lien proposé est de type participe passé également.

## **Janvier 2022 - Actudes15**

Vous avez dû vous rendre compte que certains participes passés sont présents dans le DES et que ceux que vous proposez ne sont pas toujours acceptés. La raison est simple : il serait inutile voire absurde de saisir toutes les relations des participes passés entre eux alors que les verbes correspondants sont déjà enregistrés comme synonymes.

Par exemple, la proposition consistant à ajouter visé comme synonyme de concerné n'a pas été retenue : concerné est bien dans le DES et est synonyme de en cause et en question. Visé n'est pas dans le DES car la proposition de synonymie concerne un autre participe passé. De plus les verbes correspondants, concerner et viser sont synonymes. Une autre proposition demandant d'ajouter causé comme synonyme de provoqué a été refusée pour la même raison (provoqué est dans le DES car il est antonyme de spontané, adjectif). Dans ce cas également les verbes provoquer et causer sont bien synonymes.

Une proposition de synonymie concernant un participe passé est acceptée si le synonyme est un adjectif. Par exemple, nous avons prisé qui est synonyme de populaire, adjectif ou bien enveloppé synonyme de dodu. Bien sur, certaines liaisons peuvent avoir été saisies alors qu'elles ne concernent que des participes passés. N'hésitez pas à nous en faire part dans [l'interface de propositions](#).

## **Les variantes féminines et masculines**

### **Avril 2022 - Actudes16**

Nous avons décidé il y a quelques années d'ajouter les variantes féminines en ajoutant un "-", par exemple administrateur-trice, agent-e, amateur-trice, ami-e, artisan-ane, etc. Or au fil du temps, nous nous sommes rendu compte que cela posait deux problèmes importants :

- les utilisateurs effectuant une recherche, par exemple, "policière" sans utiliser la complétion automatique, ne trouvait aucun résultat !
- les fichiers du DES utilisés dans le cadre du contrat de prêt n'étaient pas adaptés à un traitement automatique par programme : en effet comment différencier un mot normalement constitué d'un tiret, par exemple "sous-estimer" ou "avant-scène" d'un mot représentant une variante ?

Nous avons donc décidé à partir de février de cette année de modifier progressivement ces entrées avec variantes en ne gardant que la forme canonique (telle qu'elle est présentée dans les dictionnaires) et d'ajouter un lien vers la variante de façon à ce que la saisie de la variante renvoie bien vers la forme canonique retenue (policière vers policier). Il serait bien sûr

approprié d'afficher également la variante qui est saisie et pas seulement l'entrée retenue dans sa forme canonique. Malheureusement, cela demande un peu plus de développement qui sera envisagé lors de la création d'une future interface.

## Le traitement des synonymes isolés

### Octobre 2020 - Actudes10

En parallèle du traitement mensuel des propositions des utilisateurs, un important travail de gestion des synonymes isolés est en cours. Les **synonymes isolés** d'une vedette sont les mots **uniquement reliés à la vedette et à aucun autre synonyme de cette dernière**. En effet, sur l'ensemble des 50.000 entrées du DES, plus de 70 % ont de 1 à 25 liens synonymiques isolés. Nous avons depuis juin dernier traité 82 entrées qui ont plus de 50 synonymes dont 5 isolés. Un traitement automatique nous a permis de proposer pour les  $82 \times 5 = 410$  paires jusqu'à 5 synonymes de la vedette selon un calcul donné provenant d'une adaptation des liens manquants probables (et détaillé dans une [publication parue depuis sur HAL](#)). Ensuite, une vérification manuelle a eu lieu sur ces 2050 triplets ( $82 \times 5 \times 5$ ).

Voici quelques exemples de synonymes proposés et validés :

<b>Vedette</b>	<b>Synonyme isolé</b>	<b>Synonyme proposé</b>
abattement	lourdeur	molesse
apparence	superficialité	tape-à-l'oeil
attache	bouton	fermoir
heureux	optimiste	gai
malpropre	négligé	crasseux
protéger	couver	préserver
réduire	rationner	limiter
sauvage	incivilisé	primitif
usé	amoindri	émoussé
vain	sans effet	inefficace

Ces exemples sont déjà présents dans le DES.

Une 2<sup>de</sup> vague de traitement a débuté en juillet dernier portant sur 103 vedettes qui ont 4 synonymes isolés : pour chacun, de 5 à 8 synonymes sont proposés, à valider ou pas. Cela correspond au minimum à  $103 \times 4 \times 5 = 2060$  triplets à traiter ... L'état d'avancement global est consultable dans le tableau de la rubrique "comment le DES est mis à jour ?" sur la page de présentation du DES.

### Février 2021 - Actudes11

En complément de ce qui a été précisé dans la dernière lettre, un autre point que je souhaite exprimer concerne la nature des liaisons

synonymiques. En effet, qu'elles soient proposées par les utilisateurs ou calculées par programme, elles doivent bien sûr être présentes dans les dictionnaires mais également ne pas être trop spécifiques, vieilles ou rares. En effet, les mots trop spécifiques, vieillis ou rares vont très souvent être reliés à un seul autre mot plus courant et ils sont très souvent des synonymes isolés.

Or l'étude sur les synonymes isolés détaillée dans la lettre d'actualité précédente nous montre que 35.000 entrées (70% du total) ont entre 1 et 25 synonymes isolés. Plus précisément 58,9 % des entrées ont 1 synonyme isolé, 21,35% ont 2 synonymes isolés et respectivement 8,94%, 4,33% et 2,36 % ont 3,4 et 5 synonymes isolés.

Si dans certains cas, le synonyme isolé d'une vedette peut être rattaché à un autre synonyme de cette vedette, il existe aussi des cas où ce n'est pas possible car le sens est trop spécifique, vieilli ou rare. Or, il est important de ne pas avoir une trop grande proportion de synonymes isolés car les graphes sont "pauvres" et les calculs qui en découlent, que ce soit l'espace sémantique ou d'autres méthodes de regroupement de sens, ne sont pas optimaux.

Sur la seconde série d'étude des synonymes isolés portant sur une partie des entrées possédant 4 synonymes isolés (voir la lettre d'actualité précédente) où 406 synonymes isolés de 102 vedettes ont été traités, 225 ont été reliés à au moins un des cinq synonymes proposés par calcul (soit 55%) et 94 ont été supprimés, dû entre autres à leur caractère spécifique, vieilli ou rare. Quant aux 87 restants, ils ont trouvé un synonyme ailleurs que dans la liste proposée.

## **Les relations litigieuses**

### **Février 2021 - Actudes11**

Ces relations, comme cela est expliqué sur [l'interface du DES](#) ( lien : "Avant toute consultation, merci de lire cet avertissement") sont particulières dans le sens où un des deux termes de la relation a une connotation familière voire péjorative. Ces relations introduites assez récemment (par rapport à la date de création du projet dans les années 1990) font partie de la langue française car elles sont référencées comme telles dans les dictionnaires, en particulier le CNRTL et le Grand Robert.

Cela dit, elles représente 0,08% de l'ensemble des relations synonymiques (environ 187 sur un total de 210.366). Nous avons néanmoins corrigé certaines de ses liaisons suite à quelques retours et aussi parce que le contexte actuel nous amène à en revoir certaines en particulier liées aux personnes. L'objectif du projet de recherche du DES au départ n'était pas lié à ces liaisons particulières donc nous ne souhaitons pas que nos intentions soient interprétées comme volontairement discriminantes.

Enfin, ces liaisons sont maintenant accompagnées d'un point d'exclamation "!" plus approprié ( et non plus d'une étoile ).

## Les expressions figées

A partir du dictionnaire [Le Robert Expressions](#), un certain nombre d'expressions figées ont été insérées dans le DES. Ce n'est qu'un début puisque nous en avons saisi un peu plus d'une centaine. L'idée n'est pas seulement d'associer l'expression à une entrée mais également de vérifier si elle ne peut pas être associée aux synonymes de cette entrée. Cela contribue à l'obtention d'un graphe plus fourni et évite les synonymes isolés qui rendent les calculs sur les graphes moins performants.

En voici quelques unes :

### **expression 1**

entre deux portes  
frapper à toutes les portes  
frapper à toutes les portes  
à la portée de  
être en possession de  
tourner autour du pot  
tourner autour du pot  
prendre la poudre d'escampette  
se manier le pot  
tant que faire se peut  
faire machine arrière  
seconde main  
se frotter les mains  
se frotter les mains  
reculer pour mieux sauter  
ferme comme un roc  
ferme comme un roc  
sans répit  
se manier la rondelle  
se manier le pot  
se manier le train  
sur les rotules  
tirer son chapeau  
pas pour un royaume

### **expression2**

rapidement  
solliciter  
demander  
accessible  
posséder  
hésiter  
tergiverser  
filer  
se dépêcher  
dans la mesure du possible  
renoncer  
occasion  
se réjouir  
se féliciter  
attendre  
inébranlable  
inflexible  
sans cesse  
se dépêcher  
se dépêcher  
se dépêcher  
exténué  
admirer  
en aucun cas



## Les outils connexes

### Le graphe d'adjacence

#### Avril 2022 -Actudes16

Un graphe d'adjacence va représenter sous forme de **sommets** les mots, synonymes ou vedettes, et sous forme de traits appelées **arêtes** les liaisons synonymiques entre tel ou tel mot ou sommet. Par exemple, nous avons ci-dessous une illustration de graphes d'adjacence 2D et 3D portant sur *curieux*.





## Choisir un type d'affichage :

2D

2D Clusters

3D

3D Clusters

Fichiers à utiliser

fichiers Février 2022

Vedettes à saisir séparées par ";" :

curieux

Générer aléatoirement

n\_clicks : "0"

Enlever l'affichage de la(les) vedette(s)

Aide

Enlever l'affichage des arêtes de la(les) vedette(s)

Suppression des sommets qui ont moins de N liens :



Aide

Méthode de regroupement choisie :

Multilevel

? Multilevel

? Infomap

? Spinglass

Tout d'abord le graphe est affichable en 2D et 3D avec les vedettes en rouge et les synonymes en bleu. Plusieurs versions des données sont disponibles : par défaut celle de février 2022 est utilisée. Nous pouvons soit saisir la ou les vedettes (séparées par des ";"), soit cliquer sur le bouton "générer aléatoirement" qui va afficher une vedette parmi toutes celles ayant plus de 50 synonymes. Ensuite, nous avons deux boutons, l'un qui permet l'affichage de la vedette ou pas, ceci afin d'obtenir plus de clarté, l'autre qui permet d'afficher les arêtes de la vedette vers ses synonymes. Ces dernières sont par défaut affichés en pointillés. Enfin le dernier paramètre permet de supprimer les synonymes les plus éloignés du centre c'est-à-dire ceux qui

ont le moins de liens, ce nombre est représenté à côté de l'intitulé de chaque sommet.

L'interface a été améliorée dernièrement afin de calculer des regroupements de synonymes en fonction de trois méthodes : **Multilevel**, **Infomap** et **Spinglass**. Des boutons d'aide ont été ajoutés pour faciliter la compréhension des paramètres en particulier les algorithmes de regroupement.

Un menu en haut à droite de la forme suivante :



est disponible pour interagir sur le graphique. Ce menu assez intuitif est détaillé sur cette page.

Le lien pour accéder à cette interface est : <https://crisco3.unicaen.fr/desgraphe/>

## L'espace sémantique

### **Juin 2018 - Actudes3**

Une nouvelle interface de l'espace sémantique a été mise en ligne avec une ergonomie épurée et en l'illustrant avec le mot CURIEUX. A l'instar de l'interface de google maps (<https://maps.google.fr/>), le nouvel espace sémantique affiche sur un maximum d'espace et en position centrale le graphique. Les paramètres modifiables sont répartis tantôt à gauche - un clic sur "listes" fait apparaître 2 fenêtres à gauche, l'une contenant la liste des synonymes, l'autre celle des cliques, fenêtres qui disparaissent lorsqu'on clique une 2<sup>nd</sup>e fois sur "listes" - tantôt sur les axes - changement des axes : de 1 à 6- tantôt à droite avec l'aide (?), un bouton "options", et des boutons de zoom et recadrage.

[Accès à l'espace sémantique](#)

## **Les mots étudiés**

### **mutin, facétieux, factieux, mdr et lol**

#### **Avril 2018 - Actudes2**

Dans les propositions traitées en janvier, *facétieux* a été proposé comme synonyme de *mutin* avec un commentaire "en correction d'une erreur probable : *facétieux* au lieu de *factieux*". Eh bien pas vraiment, car *mutin* est bien synonyme également de *factieux*.

*Mutin* veut dire à la fois “*Qui a un caractère insoumis, rebelle, qui est porté à la révolte*” et “*Qui a un caractère espiègle, malicieux, vif ; qui est d’humeur badine, taquine*”. Dans le premier cas, il est à la fois substantif et adjectif, dans le second uniquement adjectif.

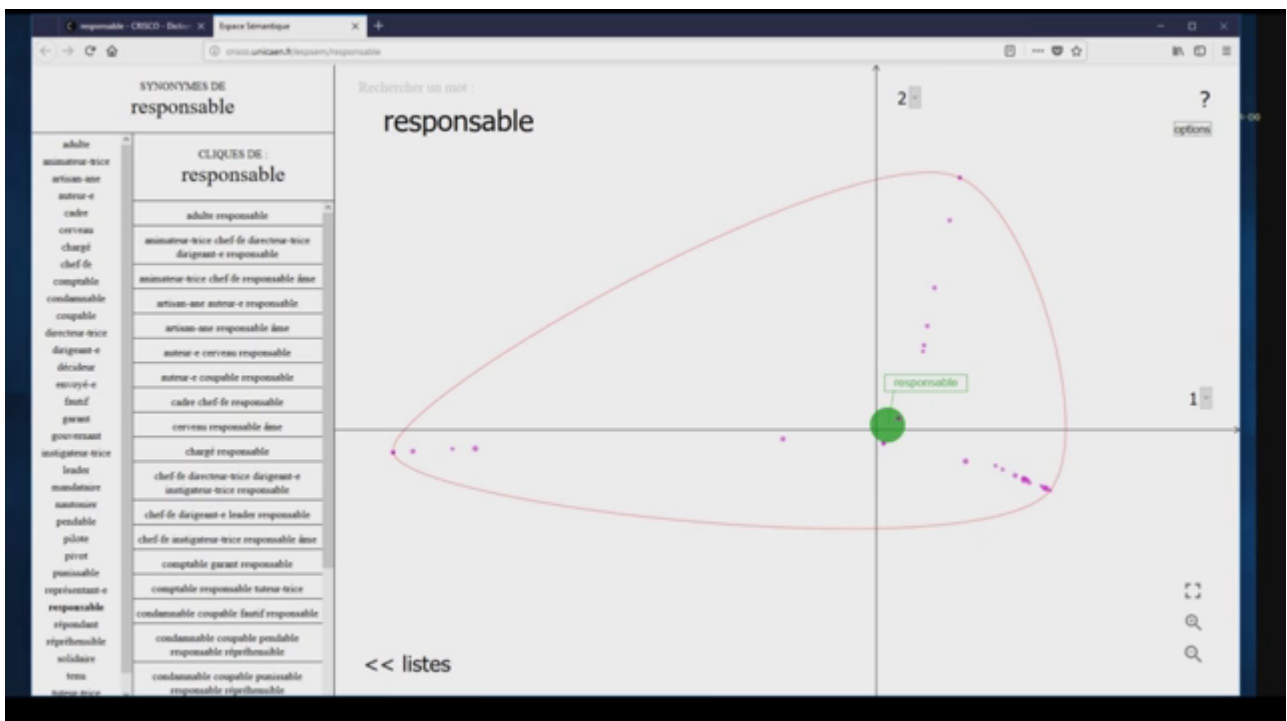
*Facétieux* est un adjectif qui décrit une personne qui “*fait ou dit des facéties, qui est naturellement ou passagèrement portée à plaisanter*”. Cela correspond bien à un des sens de *mutin* comme expliqué ci-dessus. Ceci dit en même temps, *mutin* reste également synonyme de *factieux* qui, en tant que substantif signifie “*Celui, celle qui appartient à une faction, qui a une activité subversive ... (Quasi-)synon. agitateur, insurgé, rebelle*”.

Introduction de nouveaux mots : *mdr* et *lol*. Eh oui, certains d’entre nous ne sont peut-être pas habitués à ces nouveaux mots qui font leur apparition dans les dictionnaires. Cela fait partie de l’évolution du langage et il est normal de les insérer dans le DES. Ils sont synonymes entre eux et avec *mort de rire, tordu de rire* et *plié de rire*.

## Responsable

### Mars 2019 (n°5)

*Responsable* est le mot le plus demandé en décembre 2018 et février 2019. Nous avons donc souhaité apporter quelques éléments sur sa signification. Commençons par regarder ce que l’espace sémantique nous apporte à la fois en 2D puis en 3D en visionnant cette vidéo :



Cette vidéo nous montre comment utiliser simplement l’espace sémantique 2D et nous montre la correspondance avec l’espace sémantique en 3D (accessible sur [ce lien en version interactive](#)). Pour le mot vedette *responsable*, nous visualisons 3 groupes de sens, représentés par 3

directions différentes sur le graphique. Les deux premiers sens concerne *responsable* en tant que nom : d'un coté il est synonyme de *représentant* et de *mandataire*, de l'autre de *leader*, *chef* et *dirigeant*. Le troisième sens est associé à *responsable* en tant qu'adjectif avec *coupable* et *condamnable*.

Effectivement le premier point à remarquer est que *responsable* est à la fois utilisé comme substantif et comme adjectif et cette distinction est essentielle du point de vue sémantique plus que syntaxique. En effet un esprit mal tourné pourrait conclure que tout responsable (substantif) est coupable/condamnable (adjectif). En fait on assiste sur le plan nominal à une valorisation : *responsable* > *leader* ou à un transfert d'autorité : *responsable* > *garant* / *mandataire* ( *solidaire* indique plutôt une *coresponsabilité* entre partenaires). Inversement, nous remarquons une péjoration sur le plan adjectival *responsable* > *coupable*. Mais être *responsable* c'est aussi s'affirmer comme *adulte*, sens qui apparait avec *chargé* plus près du centre. L'étymologie du mot nous apprend que *responsable* comme *répondre* ont la même origine latine "respondere" : Quiconque est (un) responsable est capable (-able) de répondre (respons-) de ses actes. Cette origine est souvent oubliée.

(merci à Jacques François, professeur émérite au CRISCO pour ces précisions linguistiques).

## Curieux

### **Juillet 2019 (n°6)**

Peut-être vous est-il arrivé de penser que l'espace sémantique n'était pas facile à appréhender ? Eh bien, ce tutoriel est fait pour vous : il vous explique en prenant comme exemple le mot *curieux* comment procéder et déduire rapidement les différents sens du mot vedette saisi.



## Espace sémantique du Dictionnaire Électronique des Synonymes (DES)

### Visualisation de CURIEUX

UNIVERSITÉ CAEN NORMANDIE



[Lien d'accès](#)

**Novembre 2019 (n°7)**

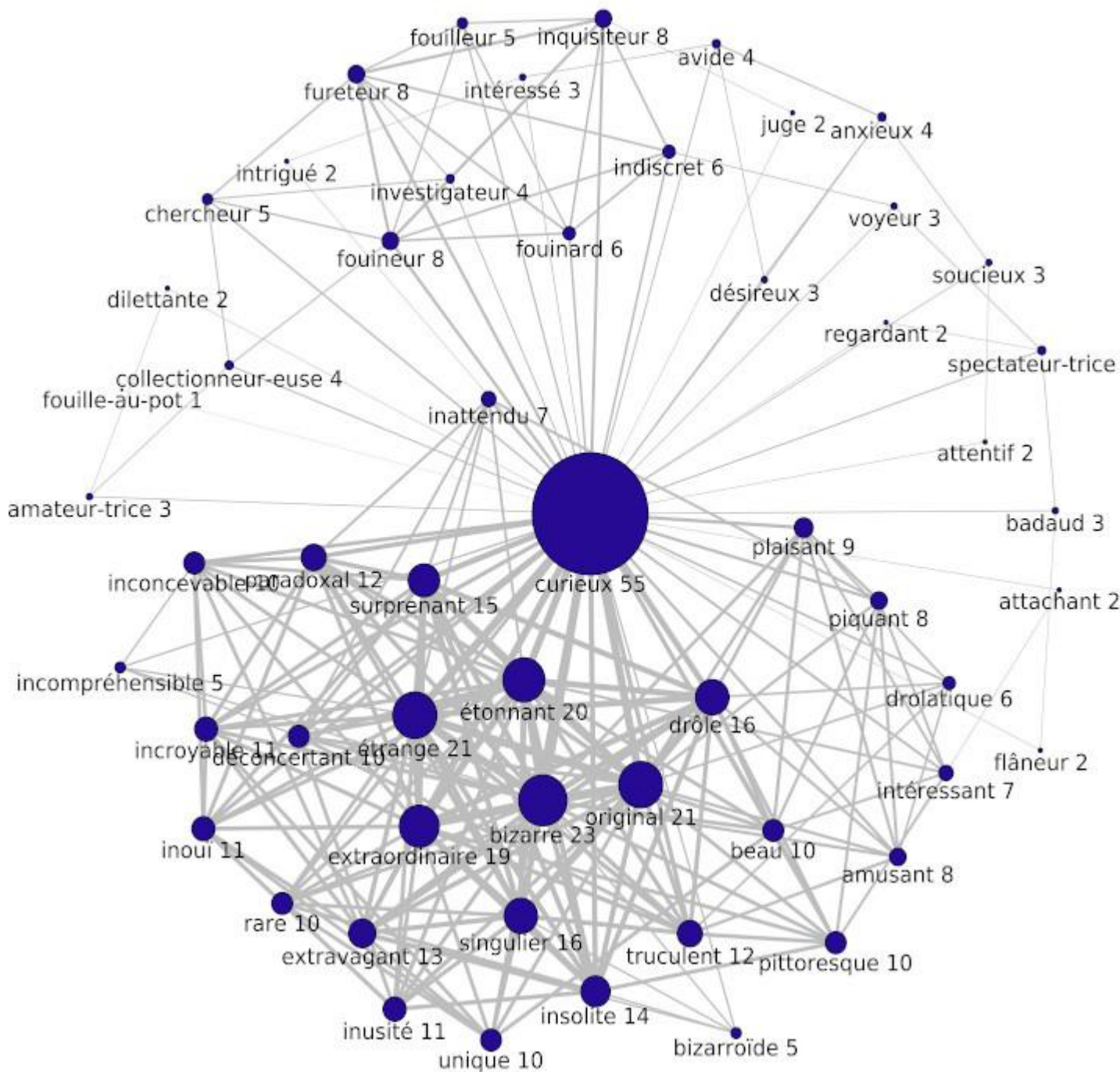
### **CURIEUX sous les projecteurs de GARGANTEXT**

Il existe, dans le même registre que l'espace sémantique, un autre outil très intéressant que je vous propose de découvrir : GARGANTEXT.

GARGANTEXT est développé par [l'Institut des Systèmes Complexes IDF, Paris](#). C'est une plate-forme en ligne qui permet l'étude de grandes données non structurées pour en saisir l'essentiel de façon dynamique.

Si nous reprenons notre exemple sur *curieux*, le graphe d'adjacence de cette vedette avec tous ses synonymes nous donnent le graphique suivant :





[Lien vers l'image](#)

Réalisé sans aucun calcul, il permet d'avoir une première idée des différents sens du mot :

- La taille des cercles des sommets est proportionnelle aux nombres de liens partant de ce sommet : le diamètre du cercle représentant curieux est bien sur le plus important puisque c'est la vedette que nous étudions.
- La taille des arcs est proportionnelle au nombre de chemins différents pour aller d'un sommet à l'autre : Plus le trait est épais, plus les 2 sommets sont liés c'est à dire qu'il existe plusieurs chemins reliant l'un à l'autre.

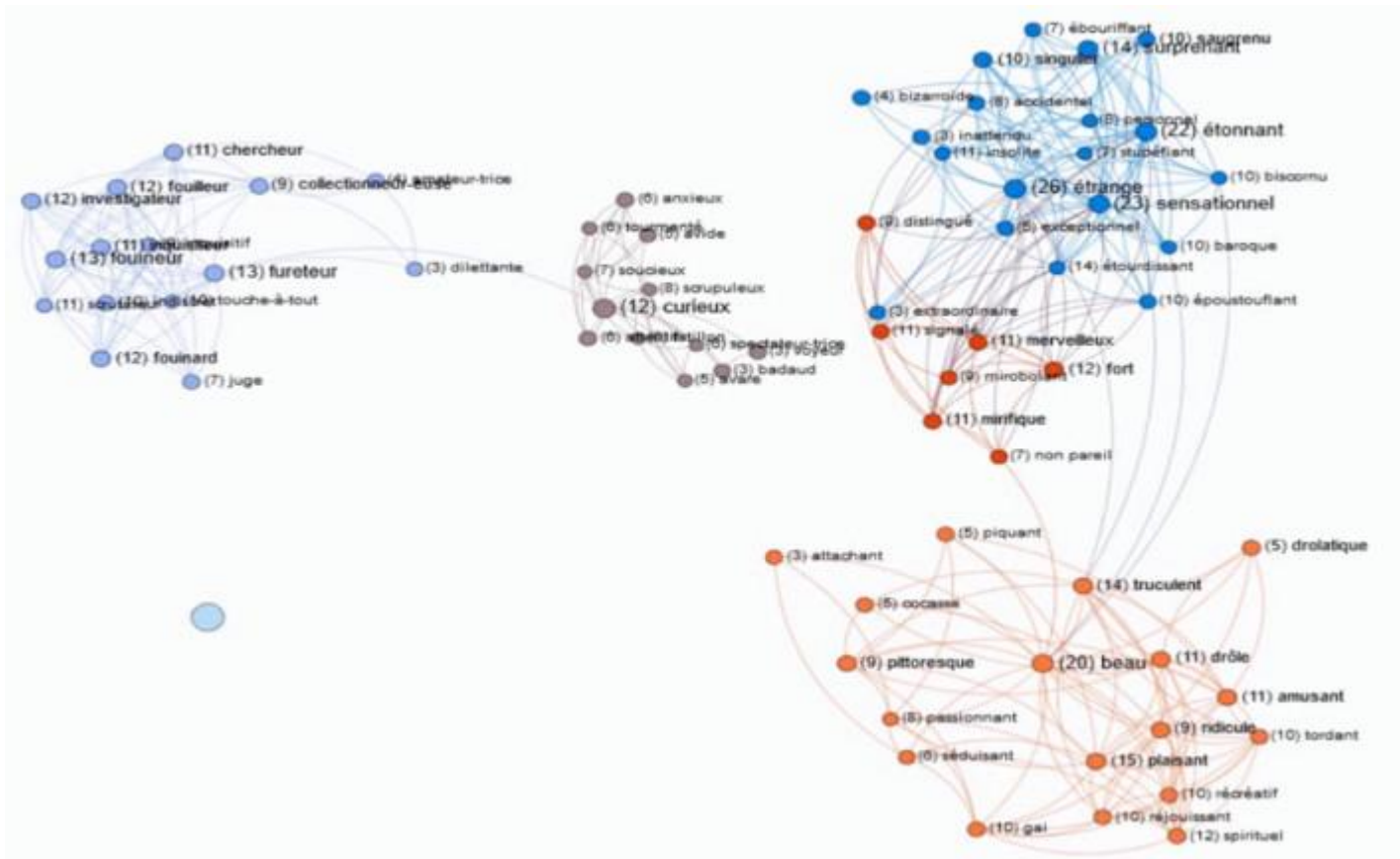
On voit bien qu'il y a beaucoup de connexions entre *étonnant*, *bizarre* *original* ... sur la droite du schéma et que les liens sont plus clairsemés sur une large partie gauche.

L'espace sémantique dans le tutoriel présenté dans la lettre n°6 nous a permis de voir quatre sens principaux de CURIEUX :

- *spectateur, badaud*
- *fureteur, fouineur*
- *original, bizarre, ...*
- *soucieux, attentif, ...*

Que pourrait nous apporter Gargantext ? Et comment cela s'articulerait avec les résultats de l'espace sémantique ?

Démonstration en image :



[Lien vers la vidéo](#)

Vous avez pu remarquer que les différents ensembles calculés correspondent en grande partie à ce que nous donne l'espace sémantique :

- *fureteur, fouineur* (en bleu ciel)
- *original, bizarre, étrange* (bleu foncé)
- *soucieux, anxieux, attentif* restent avec le cluster CURIEUX de même que *spectateur* et *badaud* (marron)

On trouve un nouveau cluster : *beau, amusant, drôle* (en rose) et si nous continuons l'algorithme de spatialisation un cluster *fort, merveilleux* (en rouge) apparaît.

Mylène Leitzelman de la société [mnemotix.com](http://mnemotix.com) a réalisé [un article sur l'importation d'un fichier tableur au format .csv dans GARGANTEXT](#).

Et puis [la documentation GARGANTEXT](#) de l'ISCPF est disponible en ligne.

GARGANTEXT est une plateforme très intéressante pour nous permettre de visualiser rapidement et de façon dynamique la polysémie d'un mot ou d'un groupe de mots et ceci en complément de l'espace sémantique du CRISCO. Les outils mathématiques utilisés dans les 2 cas sont différents ( [algorithme de Louvain](#) et [la spatialisation](#) pour GARGANTEXT, les [cliques](#) et la [réduction de dimension -AFC-](#) pour l'espace sémantique du CRISCO) et la comparaison de l'un à la lumière de l'autre enrichit les informations que nous pouvons déduire du DES (qui comprend, rappelons-le plus de 50.000 entrées et 204.000 liaisons synonymiques).

Pour en savoir plus sur GARGANTEXT : <https://gargantext.org/>

## Gagner

### Juillet 2019 (n°6)

#### **GAGNER, une variété de sens à s'y perdre, ou Qui s'y perd y gagne**

En 1930, Jean Giono a publié le dernier volume de sa trilogie sur la vie rurale en Haute-Provence sous le titre *Regain*. Ce substantif n'est plus en usage qu'accompagné d'un complément (*regain de jeunesse, de faveur*, etc. en rapport avec *regagner la jeunesse / la faveur de qqn*). Mais à l'origine, un *regain*, c'est une « *herbe qui repousse, dans une prairie naturelle ou artificielle après la première fauchaison* » (Trésor de la Langue Française) et Giono joue sur ce premier sens de *regain*, propre à l'agriculture et sur son second sens psychologique. Ce que ce premier *regain* nous suggère, c'est que *gagner* a désigné à l'origine le travail du paysan et nous en avons la confirmation avec le substantif régional et vieilli *gagnage*, « *endroit où vont paître les troupeaux* » et de là « *petite ferme* ».

De ce sens original de *gagner* (verbe d'origine germanique, plus exactement du francique, la langue des francs), il ne nous reste plus rien, sinon l'idée d'un profit, et cette idée, évidemment présente dans *gagner de l'argent / un lot / un salaire* (13° siècle, cf. *encaisser, recevoir, empocher, obtenir, recueillir*) subsiste dans des sens dérivés, p.ex. dès la même époque dans *gagner sa querelle* (sortir vainqueur d'un procès, cf. *l'emporter, vaincre, triompher*), *gagner* (orthographié *gaignier*) un ami : se faire un ami (cf. *s'attacher*), profiter d'une nouvelle amitié ou *gagner qqn à sa cause* (cf. *persuader, attirer*), dans *gagner un lieu* (16° s., cf. *atteindre, rejoindre*), *gagner du terrain* (17° s. cf. *s'étendre, se propager, se répandre*). Une voyageuse qui aborde une ville en croyant en atteindre une autre ne la gagne pas, ce n'est pas un profit pour elle et elle s'y retrouve en plein désarroi. En outre, le verbe accueille rapidement des sujets qui n'ont rien à

voir avec des personnes, p.ex. dans *le sommeil / le rire les gagne, l'épidémie gagne* dans tout le pays ou *la mer gagne sur la falaise* (à partir du 17<sup>e</sup> s., cf., *empiéter, envahir, s'emparer de*). La personne ou la chose qui tire un profit étend son emprise et c'est bien ce dernier sens d'extension qui s'applique ici au sommeil, au rire, à l'épidémie et à la mer.

### [Visualisation en 3D du graphe d'adjacence.](#)

En relation avec la vidéo sur [hal-02376412](#)

Les différents sens cités de *gagner* sont illustrés par les images de l'espace sémantique en 3D animées ci-dessous :

- [encaisser, recevoir, empocher, obtenir, recueillir](#)
- [l'emporter, vaincre, triompher](#)
- [se faire un ami \(cf. s'attacher\)](#)
- [persuader, attirer](#)
- [atteindre, rejoindre](#)
- [s'étendre, se propager, se répandre](#)
- [empiéter, envahir, s'emparer de](#)

## Travail

### **Novembre 2019 (n°7)**

De juillet à septembre 2019, *travail* a été recherché de 2800 à 4100 fois chaque mois. Il apparaît de la 30<sup>ème</sup> à la 40<sup>ème</sup> place des mots les plus recherchés. Nous vous proposons avec Nicole Le Querler, professeure émérite du CRISCO, de décrypter ce mot au travers de son étymologie.

Dans le DES, parmi les 125 cliques de synonymes du mot *travail* on trouve *accouchement, enfantement, gésine*. On dit en effet d'une femme qui accouche qu'« elle est en travail », et que dans cette situation « le travail a commencé ». De même, la salle d'accouchement est appelée *salle de travail*. Cette acception du mot *travail* est très ancienne en français et elle est liée à l'étymologie du mot : *travail* en effet est un déverbal de *travailler*, verbe attesté en français dès 1060 et issu du bas-latin *trepalium*. Un *trepalium* était une machine composée de trois pieux, qui pouvait être utilisée comme instrument de torture. Le mot est formé de *tre*, trois, et *palus*, le pieu. C'est de *palus* qu'est issu le mot français *pal*, attesté depuis 1351, qui désigne un pieu acéré à une extrémité, utilisé au Moyen-Age comme instrument de torture. À partir de *pal* est formé le verbe *empaler*. En latin classique, *trepalium* n'existe pas, mais on trouve *palus*, le poteau, et *trepalis*, adjectif signifiant « qui a trois échelas ». L'échelas est un pieu, une perche, un piquet, un tuteur, utilisé pour maintenir droits des plants de culture ou des ceps de vigne.

L'étymologie de *travail*, dans laquelle la torture est omniprésente, explique la proximité sémantique entre *travail* et *accouchement* : les douleurs de l'enfantement sont comparables à une torture. Dans la même sphère plutôt péjorative des effets de sens de *travail*, on trouve son emploi dans l'expression *travaux forcés*. Il est vrai que le premier sens de *travail*, le seul

d'ailleurs jusqu'au XVIIe siècle, était « état de celui qui souffre, qui est tourmenté ». C'est le sens qui prévaut dans l'expression *les douze travaux d'Hercule* (ici *travaux* est employé au sens de « activités pénibles et périlleuses qui apportent de la gloire »). C'est aussi la valeur qui prévaut dans ses emplois au sens de *labeur*, *besogne*, et dans ce cas la composante « gloire » de l'effet de sens est effacée. L'expression *basse besogne* accentue encore le trait péjoratif, de même que dans l'expression *basses œuvres*. De la même façon, les synonymes *corvée* ou *pensum* appartiennent à cette sphère d'effets de sens péjoratifs.

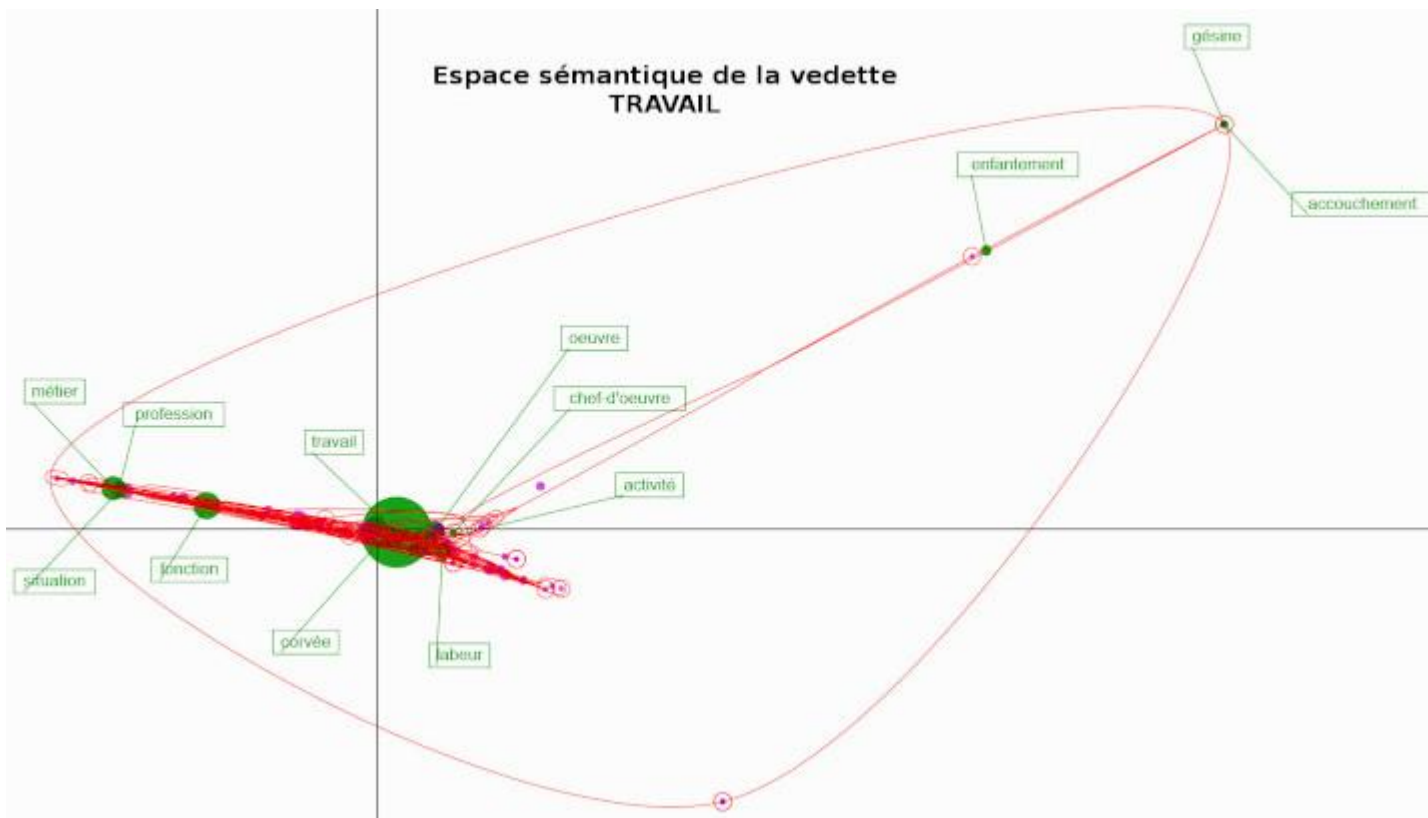
C'est aussi à partir de *trepalium* que l'on comprend le sens technique de *travail* : le mot désigne une machine permettant d'immobiliser de grands animaux, chevaux ou bœufs difficiles, pour les ferrer ou les soigner. Victor Hugo en donne un exemple intéressant : « À propos de chevaux, il paraît qu'ils sont fort méchants en Flandre, ou les Flamands fort prudents, car on ne les ferre, dans tous les villages où j'ai passé, que dans un travail des plus solides, non en chêne, mais en granit. » (Belgique, V). Dans cet emploi, le pluriel est *travails*. D'autres emplois du mot *travail* ne sont pas péjoratifs, ils sont au contraire valorisants : quand on parle du travail d'un artiste, on parle de son *œuvre*, parfois de son *chef-d'œuvre*. Par ailleurs, dans ce cas, la valeur aspectuelle du mot est une valeur résultative.

D'autres emplois encore sont neutres (ni valorisants, ni péjoratifs) : ce sont les emplois au sens de *métier*, *fonction*, *profession*, *situation* par exemple. Alors la valeur aspectuelle de travail est une valeur d'activité.

Ainsi, à partir de l'étymologie et de l'évolution historique du mot travail, on comprend l'écart sémantique entre les effets de sens péjoratifs et les effets de sens valorisants, qui paraît étonnant au premier abord. L'emploi au sens de « activités pénibles et périlleuses qui apportent de la gloire » (les douze travaux d'Hercule) fait le lien entre ces deux sphères opposées. Et les emplois neutres (ni péjoratifs, ni valorisants) n'appartiennent à aucune des deux sphères, mais ils sont extrêmement fréquents.

Sources : Grand Robert de la langue française, Petit Larousse, Robert historique de la langue française, CNTRL (Centre National de Ressources Textuelles et Lexicales), Dictionnaire latin-français Gaffiot.

*travail* possède 88 synonymes dans le DES. L'espace sémantique se présente ainsi :



### [Lien vers l'image](#)

On observe les trois grandes “directions” qui représentent les principaux sens de *travail* cités ci-dessus : en haut à droite avec *gésine*, *accouchement*, *enfantement* à gauche *métier*, *profession*, *situation*, *fonction* (sens le plus commun) et plutôt vers le centre, la notion de *corvée*, *labour* et également *d'oeuvre* et *chef d'oeuvre*.

Plus un synonyme est situé à proximité de la vedette sur le graphique de l'espace sémantique plus il est courant et assez souvent utilisé. Les synonymes en périphérie sont plus spécifiques et/ou moins courants.

Une [version dynamique en 3D](#) nous apporte les mêmes informations que cette version 2D, ce qui n'est pas toujours le cas suivant les vedettes étudiées (voir à ce sujet le tutoriel sur *curieux* pour lequel la 3<sup>ème</sup> dimension nous apporte des informations supplémentaires)

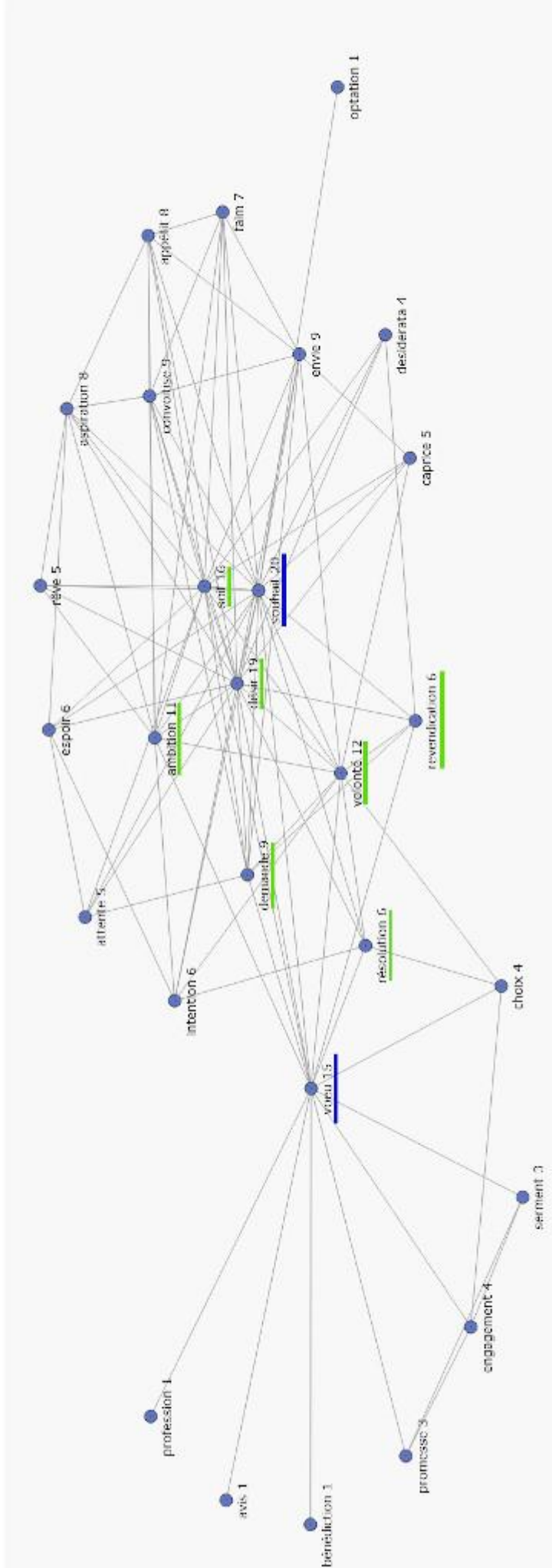
## [Les termes utilisés lors des voeux du nouvel an](#)

### **Janvier 2022 - Actudes15**

Une nouvelle année qui commence est l'occasion d'envoyer ses voeux de bonheur, réussite, santé ... à ses proches et dans son entourage professionnel. C'est aussi une période où certaines personnes en profitent pour faire un bilan de l'année passée et de se projeter au travers de projets pour la nouvelle année. Si nous partons des mots *vœu* et *souhait*, nous avons au total 27 synonymes (20 pour *souhait*, 15 pour *voeu* avec 7 synonymes en commun: *ambition*, *demande*, *désir*, *résolution*, *revendication*, *soif*, *volonté*)

répartis tel que présenté dans la figure ci-dessous (cliquer sur l'image pour l'agrandir). Chaque sommet du graphe représente un mot et chaque trait reliant deux sommets (appelé arête) signifie que les deux mots sont synonymes. Le numéro à coté de l'intitulé du sommet représente son nombre d'arêtes, donc le nombre de ses synonymes.

Graphe de VOEU;SOUHAIT (27 synonymes au total - 28 sommets affichés)





## [Lien vers l'image](#)

Soulignés en bleu, les deux vedettes et en vert les sept synonymes communs. A gauche les synonymes de *vœu* uniquement : *bénédictio*, *promesse*, *engagement* et à droite ceux synonymes de *souhait* uniquement : *caprice*, *envie*, *convoitise*, *aspiration*. Il y a également *appétit* et *faim* qui sont certainement à supprimer : ils sont effectivement utilisés avec *souhait* mais ils ne sont pas synonymes.

## **Bien**

### **Mars 2020 - Actudes8**

#### **Tour d'horizon de BIEN**

En fait, le choix de ce mot date d'avant la décision de la grève et finalement arrive à propos : qu'entend t-on par *bien* ? son contraire (ou plutôt un de ses antonymes *mal*) ? Cet adjectif, substantif et adverbe très polysémique est utilisé pour désigner une infinité de sens. Mais pour plus de précisions linguistiques, Jacques François nous a concocté une synthèse avec les exemples principaux. Pour ceux qui veulent approfondir, cet [excellent article de Robert Martin](#) devrait vous satisfaire.

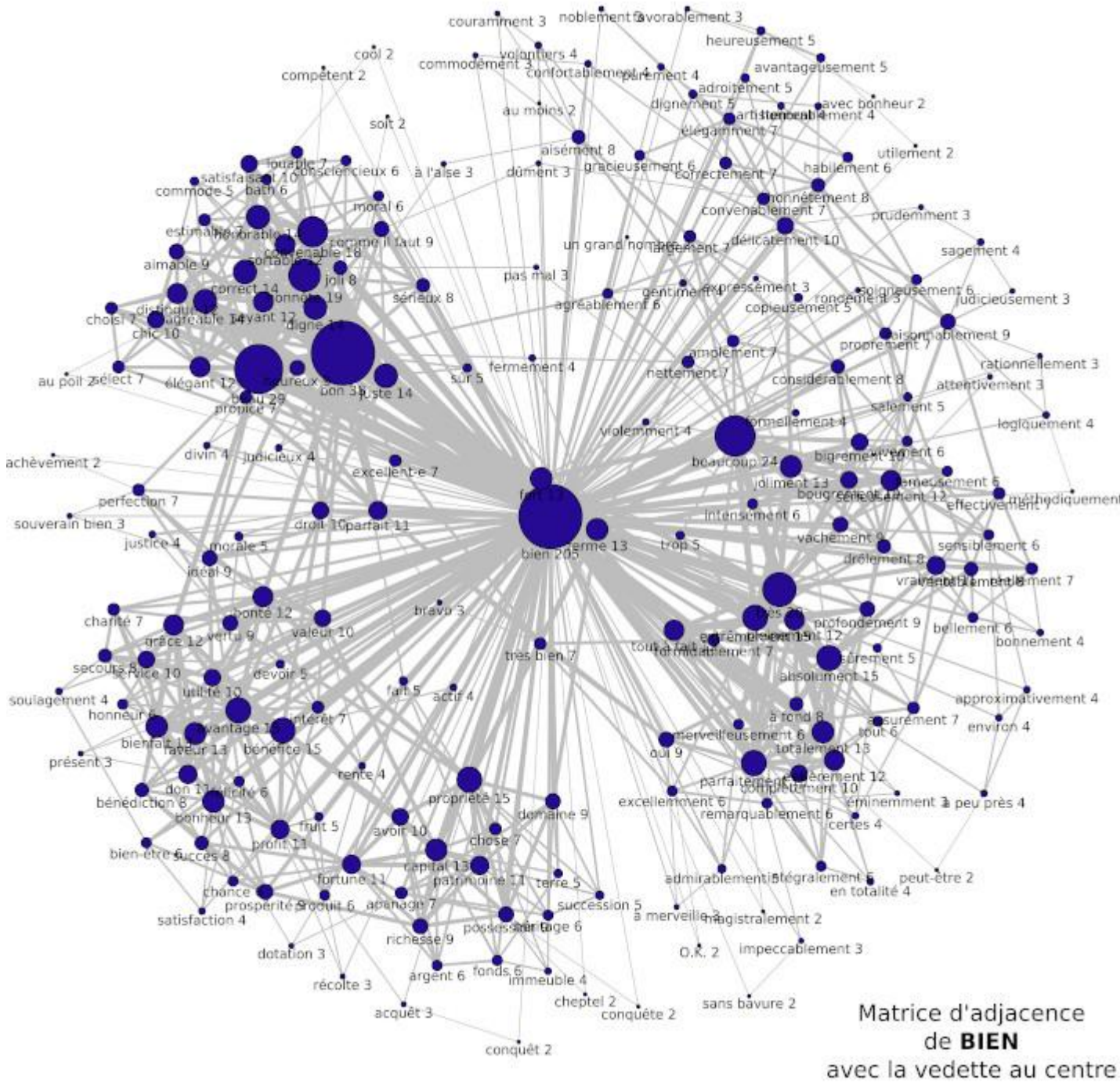
*Bien* est une des entrées dans le DES avec le plus grand nombre de synonymes : 205. C'est à la fois un adverbe, un adjectif et un substantif masculin. Un petit tour d'horizon du côté du du Trésor de la Langue Française diffusé en ligne par le CNRTL (<https://www.cnrtl.fr/lexicographie/bien>), nous apprend qu'en tant qu'adverbe *bien* est " en rapport avec certains critères d'appréciation individuels ou collectifs; d'une manière exactement adéquate à l'idée ou à l'effet attendu(e), propre à recevoir l'approbation." : j'ai cru bien faire, il a bien su dire ..., mener à bien une tâche ..., l'expression "bien à vous" et tous les emplois interjectifs : bien sûr ! très bien !

En tant que substantif masculin, le *bien* est défini dans le sens de "ce qui favorise l'équilibre, l'épanouissement d'un individu, d'une collectivité ou d'une entreprise humaine (à tous points de vue)." ou encore "Ensemble des valeurs positives fondamentales (respect de la vie et de la dignité humaine, justice, assistance mutuelle, etc.) prônées par une société donnée comme utiles à l'harmonieux développement, au progrès moral des individus, de la communauté.". Nous trouvons aussi un bien matériel : "Toute chose dont la possession, la jouissance (en fait ou en esprit) est considérée par l'Homme comme utile à la conservation, à l'expansion de son être."

Et puis cette phrase très courante : "c'est bien !" qui peut vouloir dire : c'est agréable, joli, commode, satisfaisant, juste, sérieux ... où "bien" peut-être remplacé par un adjectif.

Nous nous proposons d'aborder ces différents sens avec le graphe d'adjacence. Qu'est-ce qu'un graphe d'adjacence ? Le graphe d'adjacence d'un mot vedette va être composé de sommets qui représentent le mot vedette et ses synonymes et d'arcs non orientés représentant les relations

synonymiques entre les différents sommets. Cela donne cette première version ci-contre (cliquer sur l'image pour l'agrandir).



[Lien vers l'image](#)

Ce graphe avec les 205 synonymes de *bien* et tous les liens nous amène deux principales réflexions. D'une part nous y découvrons la richesse exceptionnelle de la polysémie de BIEN qui, comme un soleil, répand ses rayons dans toutes les directions à partir d'un centre massif (taille du cercle) : c'est finalement un objectif esthétique.

D'autre part, nous y repérons les agrégats de synonymes correspondant à des aires sémantiques et syntaxiques, notamment les trois aires de l'adjectif,

de l'adverbe et du substantif. Pour poursuivre cette réflexion, l'élagage des synonymes s'avère indispensable : nous allons perdre la symétrie générale au profit de la distinction de plus en plus claire des aires, parce que les liens entre synonymes de classes grammaticales différentes impliquent des synonymes « polyfonctionnels », susceptibles de figurer dans une phrase comme Adj/Adv et/ou N.

Quelques explications sur ce premier graphe valables également pour les suivants : la taille des cercles représentant les sommets est proportionnelle au nombre de liens partant de ce sommet. L'épaisseur des liens entre deux sommets A et B est proportionnelle au nombre de liens à supprimer dans le graphe de façon à ne plus pouvoir atteindre A à partir de B et inversement (on obtient deux graphes disjoints). Si nous reprenons le même graphe en enlevant tous les arcs concernant la vedette.

Nous obtenons le graphe ci-dessous.













également et chaleureusement pour sa relecture attentive), professeur et membre associé du CRISCO à partir de l'article de Robert Martin "Pour une approche vériconditionnelle de l'adverbe bien".

## Signe, Icône

**Mai 2021 - Actudes12**

Jacques François nous propose dans cette lettre de nous intéresser au lien unissant *signe* et *icône*. Dans son article, il présente tout d'abord les définitions de *icône*, poursuit sa réflexion en les rapprochant de *signe* puis termine par les synonymes de *signe* dans le D.E.S avec le graphe d'adjacence en l'effeuillant progressivement de façon à en extraire les différents sens.

[Télécharger l'article](#)

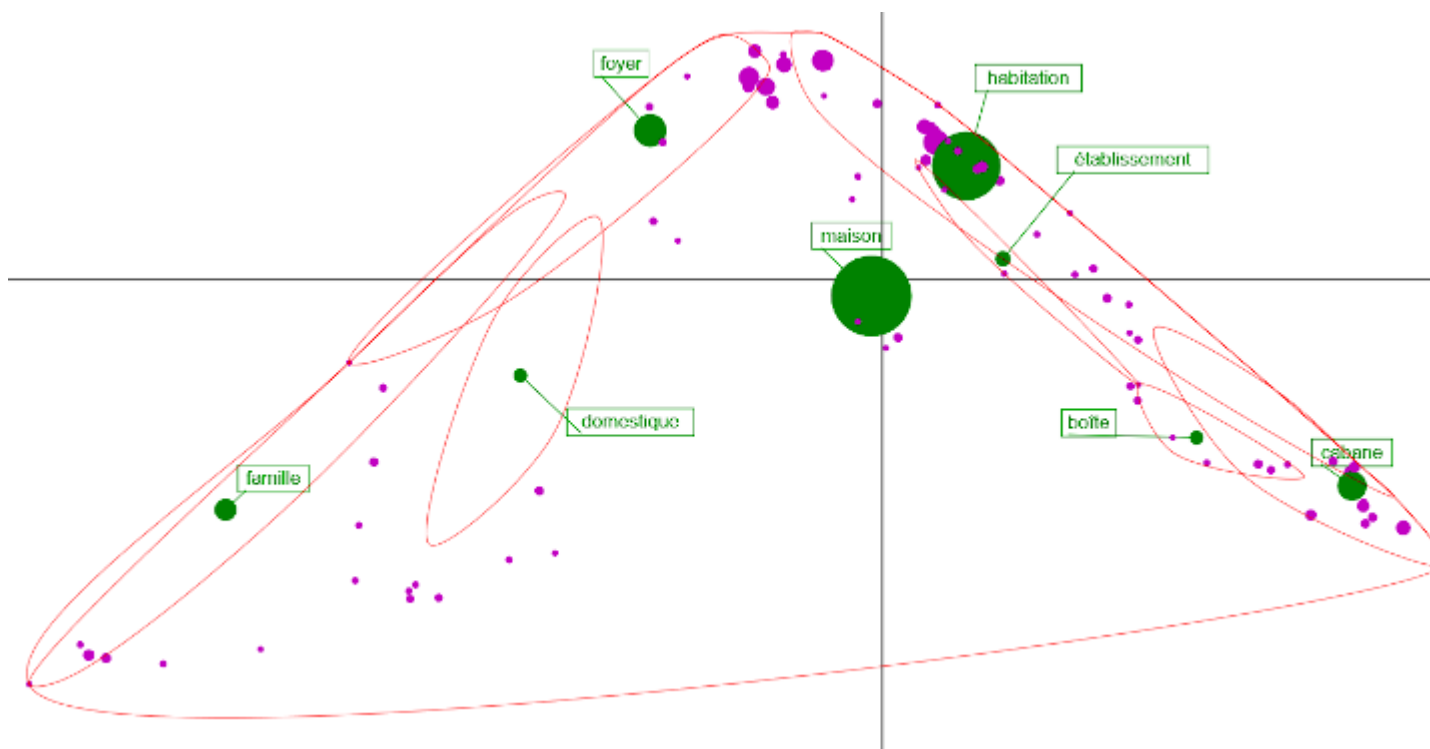
## Maison

**Octobre 2021 - Actudes14**

### **Espace sémantique de MAISON en 3D aux travers des espaces sémantiques de ses principaux synonymes**

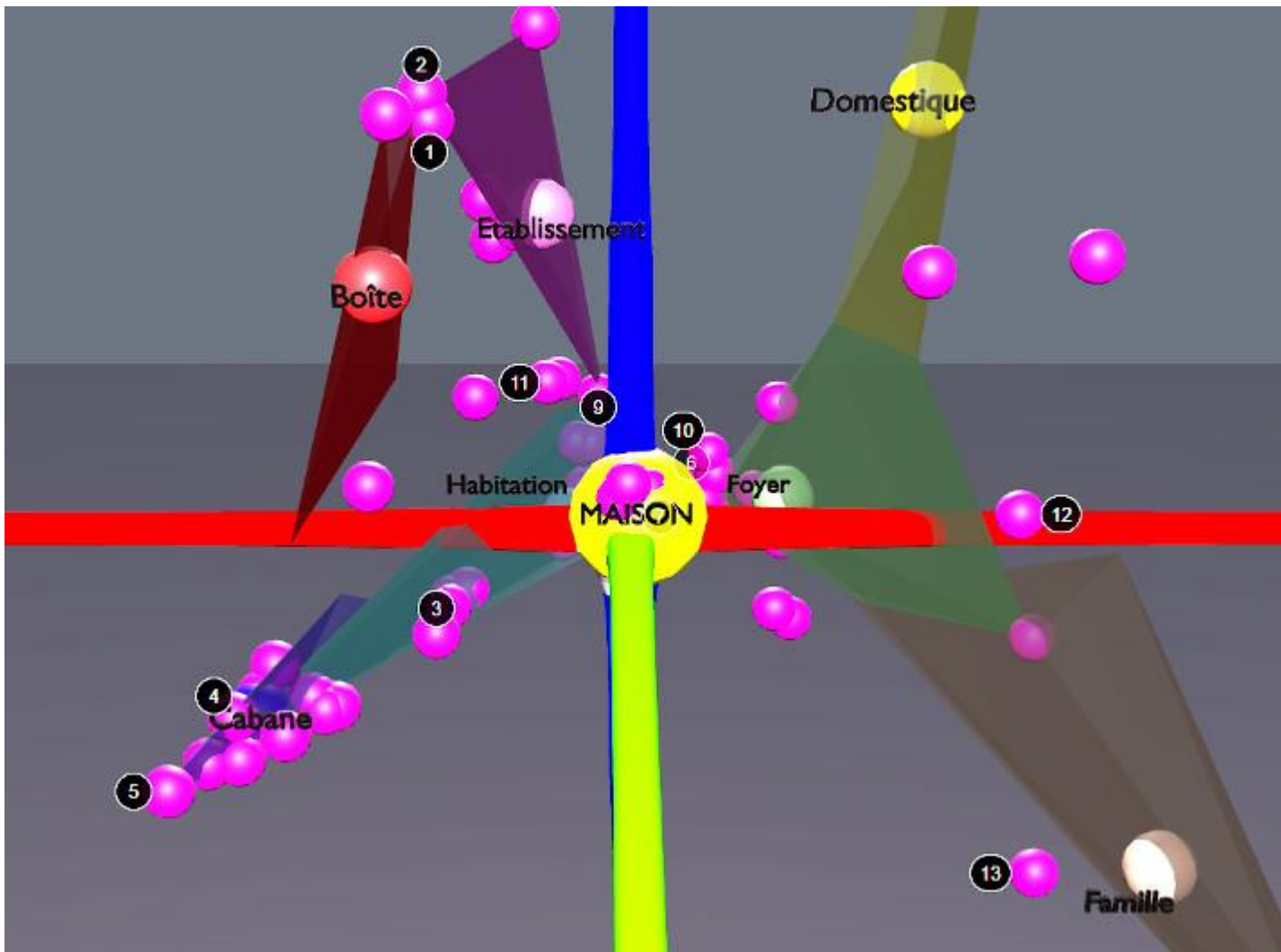
Dans l'espace sémantique d'un mot, ce dernier est toujours au centre du graphique puisqu'il est relié à tous les autres mots affichés qui sont ses synonymes. L'espace sémantique en 2D de MAISON est accessible par ce lien. Nous allons compléter cette vue 2D par une vue 3D et montrer que cette dernière permet de mieux visualiser les relations entre les synonymes.

Nous partons tout d'abord de la vue 2D en affichant les enveloppes des synonymes ayant des cercles de taille importante c'est-à-dire qui ont beaucoup de synonymes en commun avec la vedette ou en choisissant des enveloppes suffisamment larges pour être significatives. Une fois la liste de ces synonymes fixée, nous obtenons cette image :



[Lien vers l'image](#)

Intéressons-nous maintenant à la même représentation en 3D visible ci-dessous ou bien par ce lien en plein écran dans une nouvelle fenêtre:



Nous retrouvons *MAISON* au centre. Ensuite nous voyons différents polyèdres correspondant aux espaces sémantiques des principaux synonymes dont le libellé apparaît au centre du polyèdre correspondant. En partant de la vue 3D par défaut qui s'affiche (axe X en rouge de la gauche vers la droite et l'axe Z en bleu de haut en bas), nous voyons de la gauche vers le haut puis vers la droite, les espaces sémantiques (polyèdres) de *cabane*, *habitation*, *boîte*, *établissement*, *domestique*, *foyer*, *famille*.

Chacun de ces synonymes appartient à plusieurs cliques, c-a-d groupes de synonymes de *MAISON* dans lesquels ils sont tous synonymes entre eux.

Connaître ces groupes revient à réaliser une requête du genre <https://crisco4.unicaen.fr/des/synonymes/>;

Par exemple, la requête <https://crisco4.unicaen.fr/des/synonymes/maison;cabane> donne 13 cliques :

- maison;habitation → 33 cliques
- maison;boite → 4 cliques
- maison;établissement → 5 cliques
- domestique → 4 cliques
- foyer → 15 cliques
- famille → 9 cliques

Ces cliques sont représentées par des points en 3 dimensions (comme les synonymes mais non affichés pour plus de visibilité) qui délimitent le polyèdre : pour *cabane* nous avons 13 points qui délimitent son polyèdre, 33 pour *habitation*, etc. En fonction des coordonnées de ces cliques (proches les unes des autres ou éloignées), le polyèdre sera plus ou moins étiré dans les 3 dimensions. La taille va définir la portée c-a-d en quelque sorte l'importance du sens de ces synonymes dans l'espace sémantique de la vedette, MAISON.

Il est intéressant de visualiser l'emplacement de ces espaces sémantiques et leur continuité les uns par rapport aux autres : sur la gauche, *cabane* jouxte *habitation*, lui-même relié à *établissement* qui est lui-même relié à *boîte*. Puis à l'arrière *habitation* est relié à *foyer* qui se « propage » vers *domestique* en haut et *famille* en bas à droite. Cette disposition montre les liens entre les synonymes au travers de leurs espaces sémantiques créés à partir des cliques.

Nous n'avons pas par exemple *boite* à côté ou dans le prolongement de *famille* ou *domestique*.

*Boite* et *famille* (<https://crisco4.unicaen.fr/des/synonymes/bo%C3%AEte;famille>) ont 3 synonymes en commun : *école*, *maison*, *société* et dans le cadre de l'espace sémantique de *MAISON*, *école* et *société* sont absents. Donc *boîte* et *famille* sont reliés par la vedette *MAISON* uniquement. Ce qui explique leurs espaces sémantiques respectifs disjoints.

De même que *cabane* et *habitation* représentent la maison en tant que *construction*, *structure physique* alors que *foyer* et *famille* représentent la *famille* en tant que *structure sociale*. Et il y a une « liaison » entre *habitation* et *foyer* puisque ce dernier fait à la fois référence à la structure physique (lieu où l'on fait du feu) et la structure sociale.

Aux extrémités des espaces sémantiques des synonymes, des numéros s'affichent. Ce sont des annotations. En cliquant dessus, le libellé apparaît. Les n° 1 et 2 sont des mots aux limites des espaces sémantiques de *boîte* et *établissement*, ce sont donc des mots communs aux 2 espaces : nous avons *entreprise* et *firme*. Dans l'espace sémantique de *cabane*, nous avons *baraque* et *chaumière* (annotations 4 et 5) et à la limite des espaces sémantiques de *cabane* et *habitation*, nous avons *bouge* (logement misérable, annotation 3).

Entre *habitation* et *établissement*, nous avons *hôtel* et *bâtiment* (annotations 9 et 11)

Et à l'arrière entre *foyer* et *habitation*, nous avons *domicile*, *résidence* et *logis* (annotations 6,7 et 10)

Entre *foyer* et *famille*, *ménage* (n° 12) apparaît. Et dans l'espace sémantique de *famille* on a (entre autres) *origine* (n° 13)

Cette vue 3D avec les espaces sémantiques permet plus facilement de repérer l'importance des principaux synonymes d'une vedette par la taille

du polyèdre correspondant et de comprendre de part leur emplacement, le passage d'un sens à un autre.

## Schéma, diagramme, graphique, graphe et carte

### Octobre 2021 - Actudes14

Les sciences du langage sont un des domaines de recherche où - à un degré variable selon les disciplines - il est fait un usage extrêmement varié de types de visualisation soit originaux, soit empruntés à des disciplines connexes (notamment en géographie linguistique, en psycholinguistique ou en neurolinguistique? C'est aussi un champ d'étude où les modélisations jouent un rôle décisif, compte tenu ... [lire la suite](#).

## Nation, Culture

### Février 2021 - Actudes11

Jacques François, membres associé du CRISCO s'est penché sur l'évolution des sens de ces 2 mots dans le temps ...



*Cliquer sur l'image*

## Les mots les plus recherchés et les mots inexistants

### Les mots les plus recherchés - quelques statistiques

#### Mars 2019 - Actudes5

Pour les mois de décembre 2018, janvier et février 2019, les 20 premiers mots les plus recherchés sont les suivants :

Rang	Mots recherchés en décembre 2018	Nombre de fois	Mots recherchés en janvier 2018	Nombre de fois	Mots recherchés en février 2018	Nombre de fois
1	responsable	60655	synonyme	39124	responsable	299128

<b>2</b>	synonyme	34305	en effet	20462	synonyme	40057
<b>3</b>	en_effet	18775	de plus	20457	en effet	21038
<b>4</b>	permettre	10390	permettre	12475	de plus	20920
<b>5</b>	de_plus	10186	proposer	9701	permettre	12363
<b>6</b>	proposer	8234	important	8336	proposer	9778
<b>7</b>	important	7813	grâce à	8322	important	8619
<b>8</b>	ainsi	6798	ainsi	7128	grâce à	7921
<b>9</b>	par_exemple	6778	par exemple	7127	ainsi	7526
<b>10</b>	grâce_à	6681	mettre en place	6942	mettre en place	7279
<b>11</b>	mettre_en_place	5862	projet	6252	par exemple	6998
<b>12</b>	utiliser	5517	utiliser	6228	utiliser	6114
<b>13</b>	présenter	5079	développer	5942	découvrir	6072
<b>14</b>	découvrir	4858	découvrir	5892	projet	6007
<b>15</b>	dire	4798	souhaiter	5678	développer	6004
<b>16</b>	développer	4791	présenter	5649	présenter	5600
<b>17</b>	faire	4768	faire	5430	faire	5361
<b>18</b>	projet	4768	en outre	5375	comprendre	5228
<b>19</b>	comprendre	4695	comprendre	5295	expérience	5158
<b>20</b>	intéressant	4488	intéressant	5181	par ailleurs	5102

Nous constatons que ce sont très souvent les mêmes mots qui sont recherchés. Plutôt qu'une recherche véritable de leur sens, ces mots d'usage quotidien sont très probablement demandés par les internautes qui cherchent à les substituer par des mots mieux ciblés. Les différences sur ces 3 mois dans ce top20 portent sur quelques mots uniquement :

- *responsable, dire, intéressant* sur décembre;
- *souhaiter, en outre, intéressant* sur janvier et
- *responsable, mettre en place, expérience, par ailleurs* en février.

Place aux images beaucoup plus parlantes : tout d'abord les images qui résument les demandes du mois (d'où sont extraits les informations des tableaux ci-dessus) :







## [Lien vers l'image](#)

Nous remarquons de très nombreuses requêtes en février de RESPONSABLE (299128 fois), ce qui se répercute dans le nuage de mots ci-dessus et cela se traduit par un affichage plus petit des mots suivants, même si leur fréquence d'appel en février est du même ordre que celle en janvier (synonyme a été demandé 40057 fois en février et 39124 en janvier). Cette forte demande était déjà présente en décembre où RESPONSABLE a été sollicité 60655 fois. Peut-être pouvons-nous en déduire que ces requêtes coïncidant avec le changement d'année sont liées à la rédaction de rapports, bilans d'activité ou de projets ...

Et sous forme d'images animés, par jour, cela nous donne :

[Lien d'accès wordcloud201901-des.gif](#)

[Lien d'accès wordcloud201902-des.gif](#)

Il est amusant de constater les nombreuses requêtes de *JOIE, PROSPÉRITÉ, MEILLEUR, BONHEUR, AGRÉABLE, SANTÉ, SÉRÉNITÉ*, etc le 2 janvier, et dans une moindre proportion le 3 janvier. Le CRISCO a donc certainement contribué à diversifier et personnaliser la rédaction des voeux de nombreux utilisateurs :) ! Il est intéressant également de découvrir que, même si globalement nous retrouvons les mêmes mots d'un mois sur l'autre, les nuages de mots journaliers nous montrent des différences d'utilisation. Ainsi nous remarquons que le 8, 16 et 20 janvier, seuls deux à trois mots sortent du lot : *RESPONSABLE, SYNONYME, DE PLUS*.

A d'autres moments, comme le 16 janvier, le 5, 18 et 27 février, de nombreuses recherches sur plusieurs dizaines de mots se traduisent par un nuage de mots plus équilibré, avec des tailles de caractères plus homogènes. De nombreux utilisateurs du DES étaient à pied d'oeuvre ces jours là ! Enfin, il est surprenant de constater le 18 février une recherche très particulière : *GENS DE SAC ET DE CORDE* qualificatif donné à des personnes suspectes, des brigands, ... est-ce pour un roman ? un article historique ? mystère !

## **Juillet 2019 - Actudes6**

### **Les 50 mots les plus recherchés dans le DES de décembre 2018 à juin 2019.**

Sur les 150.000 requêtes reçues par jour sur le DES, nous avons pu nous pencher depuis maintenant 6 mois sur les mots les plus souvent demandés. Tout d'abord par curiosité puis ensuite cela nous permet de prendre en compte ce classement dans le choix des mots que nous étudions. Dans le nuage de mots ci-dessous, nous nous sommes limités aux 50 premiers mots les plus demandés .

Sur les 1.603.320 mots demandés durant la période de décembre 2018 à juin 2019, *gagner* apparaît en 973<sup>ème</sup> position. Dans [l'article](#) paru dans le blog de la MRSH, *comprendre, entendre, compter et importer* apparaissant

respectivement aux places 18, 925, 1099 et 3714. Quant à *curieux*, il est à la place 7183.



### [Lien vers l'image](#)

On peut constater que les requêtes portent principalement sur :

- le monde du travail : *activité, travail, projet, objectif, responsable, expérience ...*
- des verbes génériques (pour éviter des répétitions) : *permettre, proposer, présenter, dire ...*
- des mots de liaisons : *en effet, par ailleurs, par exemple, ainsi, grâce à ...*

La fréquence de *synonyme* nous interpelle. En effet, ces requêtes qui proviennent majoritairement de sites identifiés nous amènent à avancer diverses hypothèses :

- ce sont des requêtes pré-enregistrées qui s'exécutent automatiquement dès le lancement du DES ou régulièrement en arrière plan sans action directe de l'utilisateur (cookies?)
- peut-être est-ce tout simplement des internautes qui n'ont pas compris qu'ils sont déjà dans le DES et qui se demandent comment obtenir des synonymes

## Novembre 2019 - Actudes7

Les mots les plus recherchés en août et en septembre




[Lien vers l'image](#)



Les deux graphiques sont réalisés avec la librairie graphique [plotlyexpress](#) de python est dynamique. Voici quelques informations pour vous permettre de naviguer au mieux :



- dans la légende, à droite, vous pouvez décider d'afficher ou pas chacune des catégories (requêtes valides, variantes ou requêtes invalides) en cliquant dessus.

Par exemple en cliquant sur requêtes invalides dans la légende, la case devient plus opaque et l'axe Y est réactualisé pour afficher au mieux les données restantes (les requêtes valides et les variantes).

- dans le menu en haut à droite, en mode zoom  (mode par défaut), vous pouvez sélectionner une partie du graphique, par exemple les jours du mois de septembre, avec le bouton gauche de la souris. Le graphe est automatiquement réactualisé sur la zone sélectionnée. Etant plus précis, les chiffres pour chacune des catégories s'affichent. Par exemple le dimanche 1er septembre, nous avons 200.000 (200k) de requêtes valides, 12k de requêtes avec des variantes et 30k de requêtes sur des mots invalides.
- le passage de la souris sur l'histogramme affiche une fenêtre avec la date, la nature des données survolées (requêtes valides, variantes ou requêtes invalides par exemple pour le premier graphique) et le nombre.

Pour revenir à la présentation de base, il suffit de cliquer sur le bouton

“autoscale” 

- les traits verticaux de couleur magenta correspondent au dimanche de chaque semaine.
- d'autres fonctions dans le menu en haut à droite et affiché ci-dessous vous propose aussi de vous déplacer  dans le graphique ou prendre une photo  . Lorsque vous passez la souris sur un des éléments du menu, le texte de la fonctionnalité correspondante s'affiche .



## Février 2021 - Actudes11

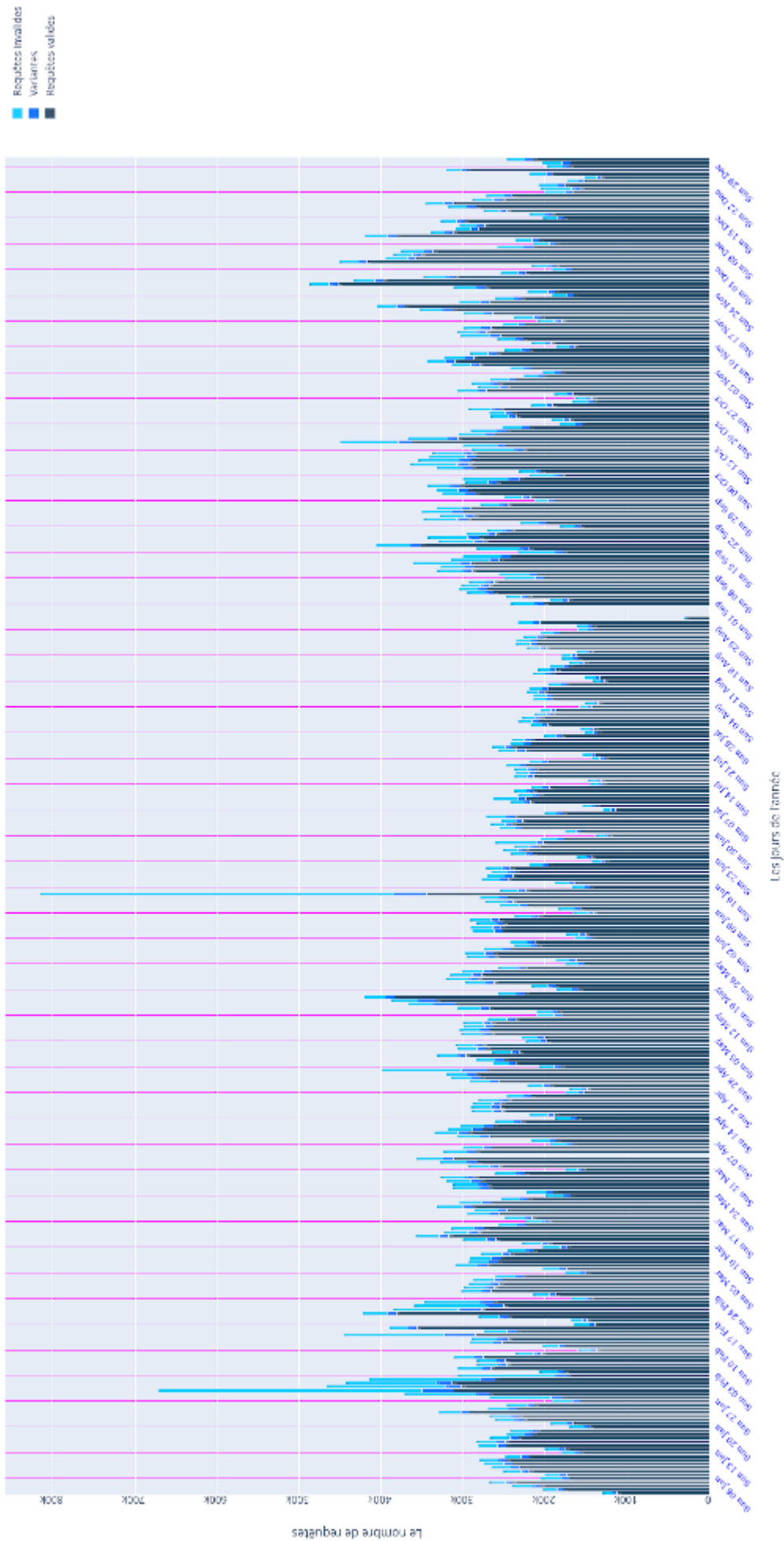
### Les mots les plus recherchés en 2019

Depuis deux ans, les requêtes sur le DES font l'objet d'un traitement mensuel de façon à extraire des logs journaliers les requêtes et les mots recherchés. Une étude complète est en cours. Voici quelques éléments en avant première.

Un premier graphique dynamique nous renseigne sur le nombre de requêtes journalières sur l'année 2019. L'aspect dynamique du graphique (comme le suivant) permet de zoomer sur certaines parties et d'obtenir plus d'informations ( voir à la fin de cette page quelques conseils pour naviguer dans le graphique).

**En 2019, chaque jour, le DES a reçu de 124.000 (7 juillet 2019) à plus de 800.000 requêtes (14 juin 2019).**

Requêtes en 2019 dans le Dictionnaire Electronique des Synonymes (DES) - CRISCO - UNICAEN



*sur l'image pour agrandir*

Ces requêtes sont classées en 3 catégories:

- en bleu foncé, les requêtes de mots valides ( = existants dans le DES)
- en bleu moyen, les requêtes des variantes ( par exemple des mots sans accents *ceder -> céder dépecher -> dépêcher* ou des erreurs courantes : *aporter -> apporter* )
- en bleu clair, les requêtes de mots invalides (qui peuvent être générés ou saisis de façon incorrecte)

Les traits verticaux de couleur magenta correspondent au dimanche de chaque semaine.

Une très grande majorité de requêtes portent sur des mots valides (traits bleu foncé). On remarque toutefois, que le vendredi 14 juin 2019 plusieurs requêtes de mots invalides sont détectées (430k) pour un nombre de requêtes valides et variantes de 379 K (340 + 39k). De même que le mercredi 30 janvier 2019, un pic de requêtes invalides (320K) a lieu de nouveau à rapport à 310K de requêtes valides et 38K de requêtes variantes.

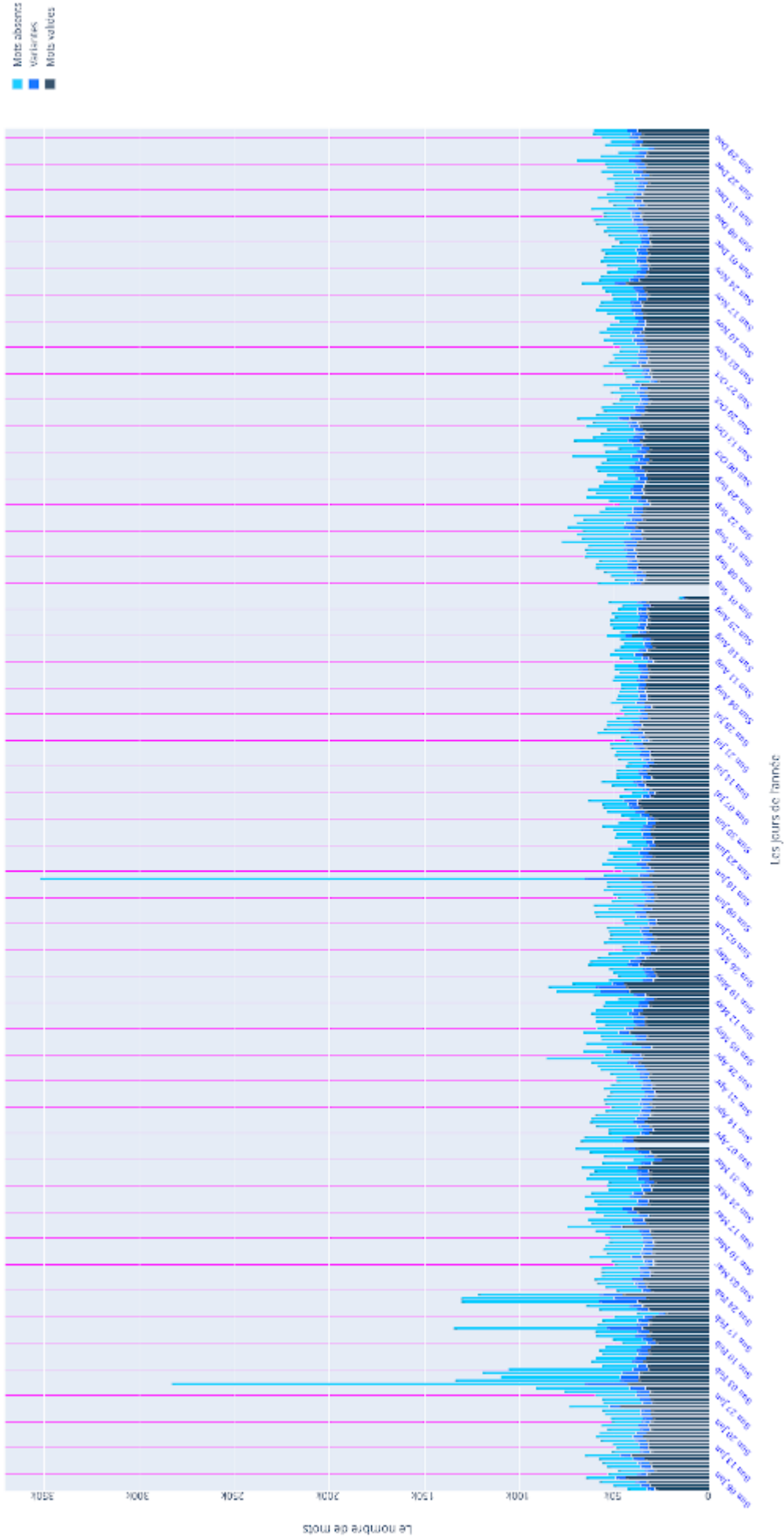
Si nous enlevons toutes les requêtes de mots invalides, nous arrivons à un nombre de requêtes valides et variantes allant de 110K+4,8K (7 juillet 2019) à plus de 460k le mercredi 27 novembre.

**A combien de mots différents correspondent toutes ces requêtes ?**

Ce second graphique va nous éclairer.



Mots les plus recherchés en 2019 dans le Dictionnaire Electronique des Synonymes (DES) - CRISCO - UNICAEN



[Cliquer](#)

sur l'image pour agrandir

**Le nombre de mots recherchés valides et différents par jour varie de 26.000 ( 23K mots valides +3.1K variantes le lundi 18 fév 2019) à plus de 60.000 (vendredi 14 juin 2019).**

Si on calcule le rapport nombre de mots valides par le nombre de requêtes valides, à la fois avec les valeurs minimales et maximales, nous en concluons qu'**un mot valide est demandé en moyenne de 4 à 7 fois par jour** (  $(23+3.1)/(110+4,8)$ ) et  $460/60$ )

De même qu'on peut s'appuyer sur le même raisonnement pour les mots absents (très majoritairement invalides) : de 2.600 requêtes invalides le 28 août à 431.307 le 14 juin 2019 , et pour les mots absents, de 2.475 le 28 août à 286.665 le 14 juin 2019. Cela donne **une moyenne de 1 à 1,5 mots invalides** par requêtes invalides. La valeur faible de cette moyenne confirme bien l'invalidité du mot demandé. Voici quelques exemples de mots absents: "Decontrctee" , "objbet de" , "itimadoulet" , "m'adébarber" , "avant-creusetto" ..

## **Février 2021 - Actudes11**

### **Mots recherchés en janvier 2021**

Les nuages de mots journaliers et le nuage mensuel sont accessibles par [ce lien](#).

Ces nuages prennent en compte les 6000 premiers mots demandés (sur environ en moyenne 50.000 mots valides demandés). Les noms de fichiers sont la forme wordcloud-des.png; par exemple pour le 4 janvier le fichier correspondant est wordcloud20210104-des.png.

Durant les premiers jours de janvier, nous avons *souhaiter, bonheur, joie, espoir, bien, réaliser* qui sont des mots très couramment utilisés pour souhaiter les voeux de début d'année. Nous retrouvons quotidiennement des mots couramment demandés : *correct, permettre, important, projet , en effet*. Certains jours il va y avoir des requêtes plus nombreuses sur un mot en particulier : par exemple *expérience* le 17 ou le 25 janvier ou *projet* le 12 et le 13.

## **Les mots inexistants**

### **Avril 2022 -Actudes16**

Les traces journalières d'accès au DES permettent non seulement de savoir quels sont les mots les plus recherchés (pour lesquels [un article](#) a déjà été publié) mais également les requêtes infructueuses. Rassurez-vous, aucune information personnelle n'est enregistrée : seule la requête l'est avec la réponse qui est retournée.

Une ébauche d'étude a été publiée dans le [carnet de la MRSH](#), la Maison de Recherche en Sciences Humaines de Caen. Elle porte sur les années 2019,

2020 et 2021 et se focalise dans un premier temps sur la proportion de requêtes infructueuses par année puis sur les mots inexistant dans le DES les plus fréquemment demandés. On peut penser, à priori, que ce sont surtout des erreurs de frappe, des mots aberrants, farfelus, insensés, qui n'ont aucun sens ... Eh bien, pas toujours et ce début d'étude pourra certainement contribuer à l'amélioration de l'interface.

Accès à l'article sur le [blog de la MRSH](#) et [sur HAL](#).

## **Les recherches particulières**

### **Juillet 2022 - Actudes17**

En dehors des 20 mots les plus couramment recherchés en **juin 2022** (*permettre, important, proposer, en effet, découvrir, projet, développer, comprendre, utiliser, présenter, ainsi, objectif, de plus, montrer, problème, travail, expliquer, réaliser, faire, créer*), nous avons constaté des particularités selon les jours. Par exemple, *loi, droit* et *parlement* le 9 et le 29, *pensée* et *sentiment* le 4,  *finesse* et *intelligence* le 7, *croyance* et *doctrine* le 10, *construction* et *bâtiment* le 15, *saveur* le 15, 25 et 28, *douleur* et *souffrance* le 3, *lumière* et *obscurité* le 6, *emploi* et *poste* le 30, ... voilà un aperçu des multiples domaines où le DES du laboratoire CRISCO #unicaen révèle ses richesses ... :) !

### **Accès aux nuages de mots**

En poussant un peu plus les recherches à partir du début d'année, nous obtenons :

#### **• Janvier 2022**

- Les 20 mots les plus demandés dans le mois : *permettre, important, mycity, en effet, proposer, projet, découvrir, correct, développer, comprendre, utiliser, de plus, faire, problème, présenter, ainsi, réaliser, travail, expérience, dire*
- Les particularités : *souhaiter, joie, doux, bonheur* (en début d'année), *entreprise, remplacement, service, vitrage, verre, double* (8 et 20 janvier), *urgence, porte, dépannage, artisan, qualité, service* (14 et 15 janvier)
- [Accès aux nuages de mots](#)

#### **• Février 2022**

- Les 20 mots les plus demandés dans le mois : *permettre, important, découvrir, proposer, en effet, projet, développer, correct, utiliser, comprendre, présenter, de plus, faire, dire, problème, ainsi, objectif, réaliser, trouver, expérience*
- Les particularités : *fanatique, ardent, violent, triste, passionné, enthousiaste, ignoble, abject, fou, exalté* (du 13 au 16)
- [Accès aux nuages de mots](#)

## • Mars 2022

- Les 20 mots les plus demandés dans le mois : *permettre, important, en effet, proposer, découvrir, de plus, développer, projet, comprendre, utiliser, ainsi, présenter, problème, correct, objectif, expérience, faire, difficulté, créer, expliquer*
- Les particularités : *effondrement, péril, menace, catastrophe (14), falaise, région (16 et 17), connaissance, diplôme, compétence, formation, travail (18), cimetièrre, citadelle, ossuaire, décomposition, corps, transformation (25 et 26), guérison (29), récompenser, offrir ( 28), défoncer, casser, discréditer, extrémiste, partisan, gauchiste, opposition (30 et 31)*
- [Accès aux nuages de mots](#)

## • Avril 2022

- Les 20 mots les plus demandés dans le mois : *permettre, important, en effet, développer, projet, proposer, de plus, ainsi, découvrir, utiliser, objectif, montrer, comprendre, présenter, expérience, correct, dire, trouver, difficile, expliquer*
- Les particularités : *hélicoptère, avion, aérien, aéronotique (1er), principe, jugement, contradiction, règle, subir (6), directeur, président, administration, directoire (12), brume, nuage, submerger, vague, noyer, couler (15 et 16), mourant, mort, cadavre, malheureux, cruel, désespoir, baton, poignard, frapper, massue, poing, torturer, violer, extraire, briser, forcer (17, 24 et 30)*
- [Accès aux nuages de mots](#)

## • Mai 2022

- Les 20 mots les plus demandés dans le mois : *péniche, permettre, important, en effet, utiliser, proposer, développer, de plus, ainsi, découvrir, projet, comprendre, présenter, montrer, expliquer, objectif, créer, trouver, intéressant, dire*
- Les particularités : *sauvage (4), roi, majesté (8), valeur, puissance, infini, absolu (10), rivière, fleuve (29)*
- [Accès aux nuages de mots](#)