

Réalisation d'un corpus sur la maltraitance

Frédérique CLAUDOT

Docteure en Droit Public et en Sciences Politiques

Université de Lorraine – UFR Médecine

CHRU de Nancy

Vivien GARCÍA

Docteur en philosophie

Université de Lorraine – UFR Médecine

Contexte

- Rapport du Haut Conseil de la Santé Publique

Dans le cadre des réflexions autour de la politique nationale de lutte contre les **maltraitements** en date de novembre 2022, le Ministre des Solidarités, de l'Autonomie et des Personnes handicapées, la Ministre déléguée chargée des Personnes handicapées et la Secrétaire d'État chargée de l'Enfance avaient souhaité que le HCSP :

- Apprécie la **production scientifique** sur ce thème
- **Fasse apparaître les lacunes dans les productions scientifiques**
- **Fasse des préconisations** en matière de lutte contre les maltraitements et **de production scientifique**

Missions

Le Haut Conseil de la santé publique (HCSP) a été créé par la loi du 9 août 2004 relative à la politique de santé publique et mis en place en 2007 (article 2).

La loi du 26 janvier 2016 de modernisation de notre système de santé a modifié les missions du HCSP par la nouvelle rédaction de l'article L.1411-4 du Code de la santé publique.

Le HCSP a pour missions de :

- 1° contribuer à l'élaboration, au suivi annuel et à l'évaluation pluriannuelle de la Stratégie nationale de santé ;
- 2° fournir aux pouvoirs publics, en lien avec les agences sanitaires, l'expertise nécessaire à la gestion des risques sanitaires ainsi qu'à la conception et à l'évaluation des politiques et stratégies de prévention et de sécurité sanitaire ;
- 3° fournir aux pouvoirs publics des réflexions prospectives et des conseils sur les questions de santé publique ;
- 4° contribuer à l'élaboration d'une politique de santé de l'enfant globale et concertée.

L'ensemble de ses travaux sont publiés sur le site internet www.hcsp.fr.

Il peut être consulté par les ministres, par les présidents des commissions compétentes du Parlement, sur toute question.

Contexte

- Temps contraint : 12 mois avec points d'étapes
- Multiplicité des institutions saisies sur le même sujet (HAS, CNSA, CNS, IGAS...)

Contexte

- La question de l'approche à adopter :
 - Par tranche d'âge
 - Par population
 - Par nature de la maltraitance
- Groupe d'expert composé majoritairement de professionnels de santé mais l'ambition était :
 - de faire ressortir les travaux issus des sciences humaines et sociales
 - de faire dialoguer les institutions (santé, médico-social, justice, police, ...)

Contexte

Volonté de rendre un travail *opérationnalisable... rapidement*

- Quelles sont les données produites en France ?
- Quelles données manquent ?

Objectif

Le GT relatif à la production scientifique devait :

- Recenser et décrire la littérature disponible sur la maltraitance envers les publics en situation de vulnérabilité
 - Approche par tranche d'âge
 - Approche par population
- Faire ressortir, plus spécifiquement, l'état de la recherche en France

Méthode

Deux travaux ont été menés :

- (l'un à partir de « Scival », développé par Elsevier, et se penchant plus spécifiquement sur les personnes âgées, réalisé par une autre équipe)
- l'autre utilisant ISTEK et LODEX, et s'intéressant à la maltraitance, toutes catégories confondues

Méthode

- Le travail mené avec ISTEK et LODEX a été effectué par 2 chercheurs, entre juillet et octobre 2023 (premier travail avec ISTEK et LODEX pour les 2 chercheurs)
- Pourquoi ISTEK et LODEX ? Car la combinaison des deux nous semblait pertinente pour l'objectif : ISTEK pour la réalisation du corpus (généraliste, multidisciplinaire, etc.), LODEX pour la catégorisation et, en partie, pour la présentation des résultats

Identification des mots clés

- FC d'abord identifié les mots-clés en français, pour les faire valider par le GT
 - Pour la traduction, nous avons utilisé le MeSH bilingue mis à disposition par l'INSERM (consultable à l'URL <https://mesh.inserm.fr/FrenchMesh/>)
 - Nous avons complété la liste des mots-clés en utilisant les arborescences du MeSH (Medical Subject Headings), thésaurus de référence dans le domaine biomédical mis à disposition par la bibliothèque nationale de médecine américaine (US NLM) et qui permet d'interroger notamment les bases de données MEDLINE/Pubmed
- Ce travail a été enrichi par une recherche de termes et traductions sur le Health Terminology/Ontology Portal (HETOP — consultable à l'URL <https://www.hetop.eu/hetop/>)

Découverte d'ISTEX

- Fin juin 2023, premiers essais avec ISTEX. Nous avons suivi les tutoriels et parcouru la documentation, mais tout ne nous était pas absolument clair (nous saisissons mal les liens entre ISTEX et LODEX, phase d'acculturation avec le vocabulaire : « enrichissement », « facette »).
- Après un mois de bricolages en tout genre, nous avons contacté l'INIST qui nous a guidé tout au long de notre démarche, à commencer par l'équation de recherche.

Histoire d'une équation de recherche (1)

- Avec l'aide de l'équipe ISTEEX, et à partir de nos mots-clés, nous avons construit une première équation de recherche.
- Un premier coup d'œil sur les résultats nous a conduit à exclure la maltraitance envers les animaux et à limiter les résultats auxquels conduisait l'utilisation de « elder* » (le bureau est hors-champ).

```
(abstract:*) AND ((title:(maltreat* mistreat* abus* violen* bully* stalk* neglect* harass*) AND (elder* child* spous* family institution* "intimate partner" emotional workplace domestic gender sexis* ethnic* racis* disability disabl* impair* resilien*)) (abstract:(maltreat* mistreat* abus* violen* bully* stalk* neglect* harass*) AND (elder* child* spous* family institution* "intimate partner" emotional workplace domestic "gender-based" sexis* ethnic* racis* disability disabl* impair* resilien*))) NOT (animal* "violet" "elderberry" "elderflower")
```

- 64 266 résultats

Premier passage à LODEX

- Importation des résultats d'ISTEX
- Utilisation d'un premier modèle clé en main pour les données Istex fournit par Pascale Viot
- ... et familiarisation avec le logiciel (et avec lodash)

Premiers défis

- Comment distinguer les travaux des équipes françaises et ceux des équipes non-française ?
- Comment catégoriser les résultats par population ?

Travaux des équipes françaises (1)

- Dans un premier temps, nous avons essayé d'utiliser le Webservice addressSplit - Découpage d'une adresse.
- Mais les résultats étaient hétérogènes : parfois le pays était renvoyé dans « house », parfois dans « country ». Et surtout il produisait plus de 24000 erreurs.

Inist-CNRS
2, rue Jean Zay
CS 10310
F-54519 Vandœuvre-lès-Nancy
France

Méthode

Une adresse postale est souvent présentée sous forme textuelle, exemple :

2, rue Jean Zay
54519 Vandœuvre-lès-Nancy
France

Elle est composée de plusieurs informations sémantiquement différentes comme le pays, la rue ou le code postal. Pour l'exemple précédent, le web service (qui s'appuie sur la bibliothèque LibPostal) retourne les informations suivantes :

```
"house": "inist-cnrs",  
"house_number": "2",  
"road": "rue jean zay cs 10310",  
"postcode": "f-54519",  
"city": "vandœuvre-lès-nancy",  
"country": "france"
```

Illustration reprise de : <https://services.istex.fr/decoupage-dune-adresse/>

Travaux des équipes françaises (2)

- Après plusieurs essais et plusieurs échanges avec l'équipe ISTEEX-LODEX, nous avons vu que les résultats les plus pertinents pour notre cas d'usage étaient obtenus par une recherche du nom du pays dans la colonne Affiliation(s)
- Limites :
 - Si un laboratoire se trouve rue Anatole France à Bruxelles, par exemple
 - Si le pays du laboratoire est absent de l'affiliation (ou s'il y a une coquille)→ nécessité de vérifier les résultats (ce qui peut-être laborieux étant donné la taille du corpus initial)

Name

Travaux fr

▶ RUN

Status : Done

[SEE LOGS](#)

Advanced mode

```
1 [assign]
2 path = value
3 value = get('value.Affiliation(s)')
4   .map(item => item[0].search(/[Ff]rance/i) !== -1)
5   .reduce((acc, current) => acc || current, false)
6   .toString()
```

Autre catégorisations

Pour les autres catégories (tranches d'âge, thématiques, etc.), là encore après plusieurs essais, nous avons adopté l'approche suivante :

- Extraire la liste de l'ensemble des mots-clés Teeft des articles
- Procéder à une catégorisation « à la main »
- À partir de ces listes de mots, créer des enrichissements correspondant à chaque catégorie de mots-clés pertinents.

Name

Prévention

▶ RUN

Status :

Done

[SEE LOGS](#)

Advanced mode

```
1 - [assign]
2   path = value
3   value = get("value.Mots-clés (teeft)").castArray().some(v => ["violence prevention", "prevention program", "injury prevention", "primary prevention", "violence prevention program", "prevention effort", "family violence prevention fund", "prevention strategy", "crime prevention", "delinquency prevention", "child abuse prevention"].includes(v)).toString()
```

Histoire d'une équation de recherche (2)

En travaillant sur nos listes de mots-clés, nous avons pris conscience qu'il y avait beaucoup de bruit dans nos résultats.

Notre première réaction a été de reproduire nos méthodes de travail habituelles → Faire un *screening* comme on le fait dans le cadre d'une revue de la littérature (en définissant des critères d'inclusion et de non-inclusion sur titre et résumé). Cependant :

- Le temps était compté et il aurait fallu faire ce travail sur 64266 résultats
- Certains types de résultats non voulus étaient récurrents (ex : textes d'astrophysique) et devaient pouvoir être filtrés par une équation plus fine.

Histoire d'une équation de recherche (3)

- Nouvelle tentative :
 - Une bonne part des résultats non voulus sont liés à l'utilisation de abus* et neglect* dans le titre (Il est pertinent d'utiliser ces termes dans le résumés, mais pas nécessairement dans le titre)
 - war et militar* et gang* sont à exclure (s'ils impliquent des actes de violences sur les personnes et les biens, ils ne correspondent pas à la commande)
 - Essai de tri des résultats en excluant certaines catégories scienceMetrix (ex : "Agricultural economics & policy")
- 11894 résultats ! Qui correspondent davantage à nos attentes.

Histoire d'une équation de recherche (3)

Mais nous ne nous sommes pas arrêtés là, car une fois notre nouveau corpus ISTEEX chargé dans LODEX, nous n'étions pas complètement satisfaits du résultat.

Pour comprendre ce qui n'allait toujours pas dans notre équation de recherche, nous avons largement utilisé la fonctionnalité « parcourir les résultats » de LODEX.

À travers cette recherche, l'un des grands avantages du couplage ISTEEX-LODEX nous est apparu : « jouer » avec les enrichissements et les facettes est aussi un moyen efficace pour juger de la pertinence d'un corpus et affiner, de manière itérative, une équation de recherche.

Histoire d'une équation de recherche (fin)

- Ce processus nous a conduit à préciser (encore) nos mots-clés et à nous faire nous prendre conscience qu'une erreur d'opérateur booléen nous était passée sous le nez. Dans notre première équation, nous recherchions par titre **ou** résumé plutôt que par titre **et** résumé.
- Avec cette modification, l'exclusion de catégories scienceMetrix était tout bonnement inutile.
- 10 062 résultat

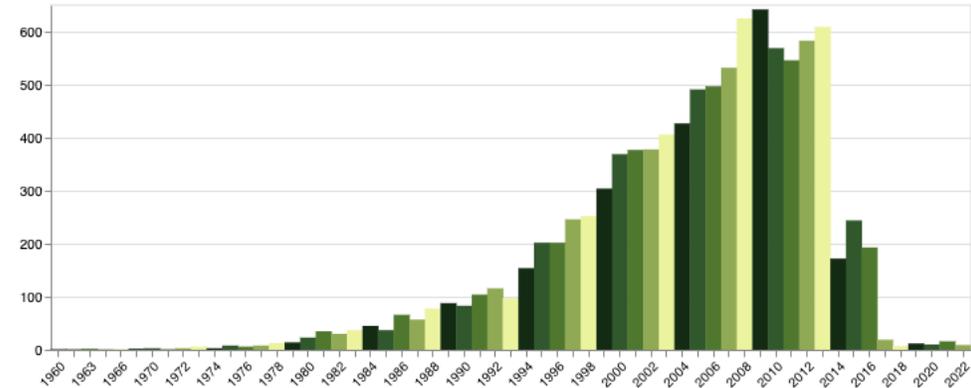
```
(abstract:*) AND ((title:(maltreat* mistreat* bully* /stalk[ei].*/ harass* violen* abus* neglect* maltrait* /harc[eè]l.*/ néglig*) NOT ("war" militar* gang* animal bacterium)) AND (abstract:(maltreat* mistreat* bully* stalk* violen* harass* maltrait* /harc[eè]l.*/ néglig*) AND (senior adolesc* elder* child* spous* family institution* "intimate partner" emotional workplace domestic "gender-based" sexis* ethnic* racis* disability disabl* impair* resilien* neglect* abus*))) NOT (genre:"abstract" genre:"brief-communication" genre:"book-reviews" genre:"editorial")
```

Affichage des résultats dans Lodex

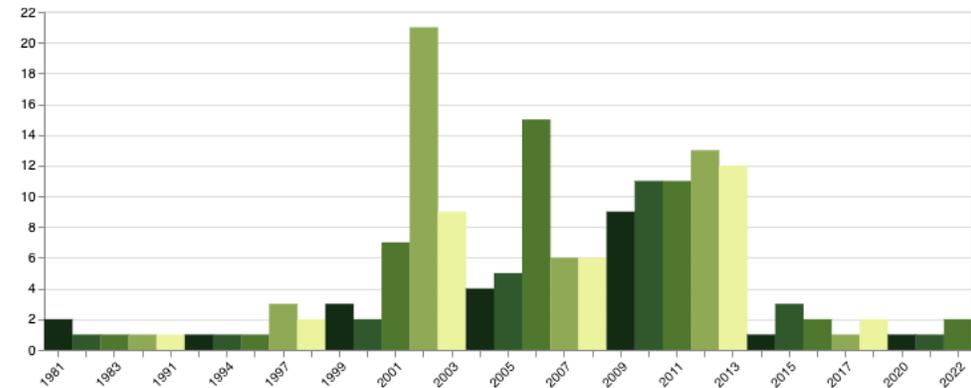
- Nous avons assez peu utilisé les capacités de LODEX ayant trait à l'affichage des résultats (ex : création d'une page web)
- Nous nous sommes contentés d'extraire les représentations graphiques et des tableaux

Quelques résultats (1) Évolution chronologique des recherches sur la maltraitance

- Publications sur le sujet depuis les années 1960 à l'international et depuis les années 1980 en France
- Croissance comparable FR-INT, avec quelques pics en France
- Baisse des publications après 2013 (à nuancer, avec le fait que les publications plus récentes n'étaient pas nécessairement référencées dans ISTE).



Nombre de publications (ordonnée) par année (abscisse) à l'international



Nombre de publications (ordonnée) par année (abscisse) en France

Quelques résultats (2) Analyse du corpus par catégories

	International	UE (hors Fr)	France
Enfants, jeunes, adolescents	4829	471	19
Familles	3725	245	15
Genre	2448	270	9
Environnements de maltraitance	674	78	5
Patients	441	69	4
Personnes âgées	316	32	3
Handicap	159	18	2

Quelques résultats (3) Maltraitance et santé publique

	International	UE (hors fr)	France
Santé publique	1408	152	5
Prévention	1078	61	4
Promotion de la santé	93	11	0

Remerciements

Nous remercions la *team* ISTEEX - LODEX :
Camille de Salabert, Pascal Cuxac
et plus particulièrement
Pascale Viot, Léo Gaillard pour leur aide, leur
disponibilité, la rapidité de leurs réponses, leur
efficacité et leur aimable accueil.