



Lyon 1

EISO
Master Epistémologie
et Ingénierie
de La Science
Ouverte

Analyse des travaux de recherches sur le handisport dans la base de données HAL



Enseignant : Rémy CAZABET



Sommaire

| | |
|---|----------|
| Sommaire | 1 |
| Introduction | 2 |
| Analyse de la requête | 3 |
| Évolution dans le temps du nombre de publications | 3 |
| Domaine de recherche - gregory | 4 |
| Réseaux d'auteurs | 5 |
| Répartition en France | 6 |
| Type d'handicap traité | 7 |
| Conclusion et ouverture | 8 |
| Annexe | 9 |

Introduction

A l'aube de notre masterclass sur le thème "sport et science", nous vous proposons un rapport qui traite d'une sous-partie de la masterclass : le handisport.

Nous avons interrogé l'archive ouverte HAL ,et extrait des métadonnées de son API; qui ont été traitées via python sur Google colab.

Les résultats seront présentés sous différentes formes tels qu'une carte, un histogramme, un diagramme camembert et un réseau pour faire une analyse assez large de la question. Il a donc été choisi de faire un travail de surface car ce sujet est assez récent dans le monde de la recherche.

A la fin de ce rapport des pistes seront dégagées pour travailler en profondeur sur ce jeu de données.

Nous avons utilisé différentes requêtes en fonction des besoins. La première requête contient même des informations qui ne nous ont pas servies. La requête final retenu pour ce travail contient 226 articles avec les données suivantes :

- Titre
- Mots clés
- Date de publication
- Adresse de la structure
- Domaines
- Auteurs

Soit, une fois rédigé sur le Google Collab :

```
response = requests.get  
("http://api.archives-ouvertes.fr/search/?q=handisport* OR paralymp* OR  
parasport*&wt=csv&rows=300&fl=title_s,keyword_s,publicationDate_s,struct  
tAddress_s,domainAllCode_s,authFullName_s")  
as_csv =pd.read_csv(io.StringIO(response.text), sep=",")  
as_csv
```

Analyse de la requête

Évolution dans le temps du nombre de publications

Le nombre de publications par année est une des premières analyses auxquelles nous avons pensé. Le code a été relativement simple à réaliser. Il y a juste eu un problème sur le nombre d'articles de l'année 2023 qui figurait sur l'année 2022, mais en précisant les dates souhaitées, cela s'est arrangé.

Nous nous attendions à retrouver un cycle quadriennal avec des pics de publications se situant l'année précédant les jeux olympiques (JO). Finalement les pics de publication sont assez légers et se retrouvent plutôt dans les 2 années qui précèdent les JO (2014, 2018, 2022).

Un autre point que nous pouvons relever est la date du premier article qui montre bien que les publications sont récentes (2002). De plus, elles restent peu nombreuses jusqu'en 2012, à savoir, moins de 5 par an. Depuis 2012, malgré une évolution légèrement discontinue, le nombre de publications augmente rapidement.

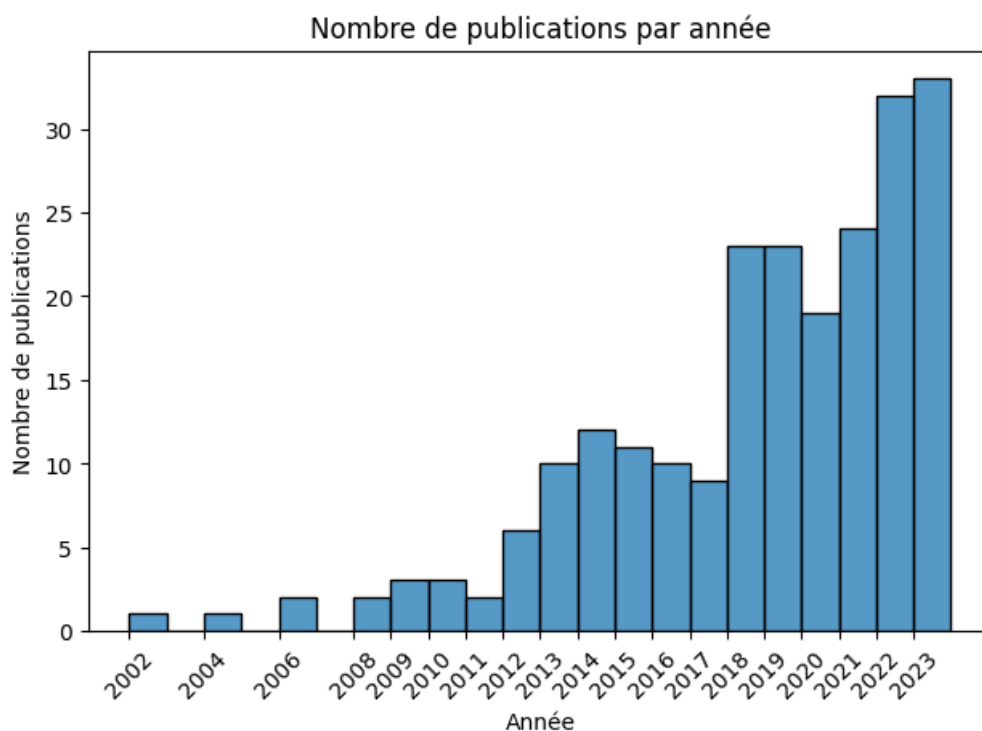


Figure 1 : Évolution du nombre de publications par année sur HAL.

Domaine de recherche

Par la suite, nous avons cherché à savoir quels sont les domaines qui ont effectué le plus de recherches sur les questions de handisport. En fonction des domaines trouvés, nous pouvons avoir une meilleure idée de l'angle pris pour les recherches sur le handisport.

Il a fallu dans un premier temps, récolter tous les domaines d'où viennent les chercheur.e.s pour chaque article. Après analyse nous avons constaté que pour chaque article il peut y avoir plusieurs chercheur.e.s, et que chaque chercheur.e.s a une spécialité. Le sunburst permet de visualiser cela et aussi de voir les domaines et les spécialités des chercheur.e.s.

Un sunburst plus précis a aussi été effectué avec le nombre d'occurrences de chaque spécialité (annexe 1). Cependant il est moins pertinent dû au nombre de répétitions des spécialités. Il y a plusieurs chercheurs par articles, ces chercheurs sont souvent issus des mêmes domaines et cela amplifie donc la répétition. Des phénomènes déjà visibles dans la Figure 2 ce retrouve donc amplifié dans l'Annexe 1.

La figure 2 montre une forte présence des SHS, avec plus de 50 % des recherches sur le handisport. Les Sciences dures quant à elles sont aussi présentes mais minoritaires. On peut donc en déduire que le handisport est surtout étudié par rapport à sa position dans la société. Et dans un second temps, ce sont des recherches visant à améliorer les performances techniques.

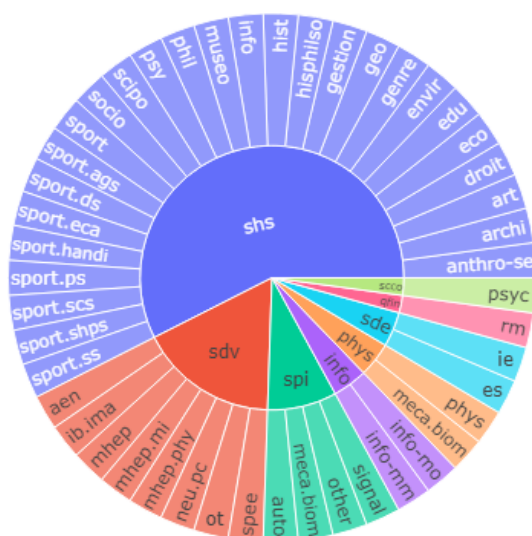


Figure 2 : Représentation des domaines et spécialités traités dans la recherche sur le handisport en France.

Réseaux d'auteurs

Une fois les domaines étudiés, nous avons cherché à savoir si les chercheur.e.s travaillaient beaucoup entre eux et comment se répartissaient leurs réseaux.

Pour cette analyse, nous avons exporté les données depuis le python de Google Collab en graphml afin de les importer dans le logiciel Gephi et représenter leur Closeness et Betweenness.

La closeness est représentée par les couleurs de rouge à bleu, et la betweenness est représentée par la taille des nœuds.

Il est alors possible d'observer que des réseaux très séparés sont présents et que les petits réseaux sont très interconnectés. Et certains aut.eur.rice. telles que Héléne Joncheray forment des ponts entre les réseaux. On peut donc émettre l'hypothèse que les chercheur.e.s des mêmes domaines travaillent souvent entre eux mais ne travaillent pas avec les autres domaines. Pour affirmer cela, il faudrait pousser la recherche plus loin et regarder le domaine de chaque chercheur.e.s dans ces groupes. C'est donc un sujet avec une forte diversité disciplinaire mais il n'y a aucune collaboration entre ces différentes disciplines.

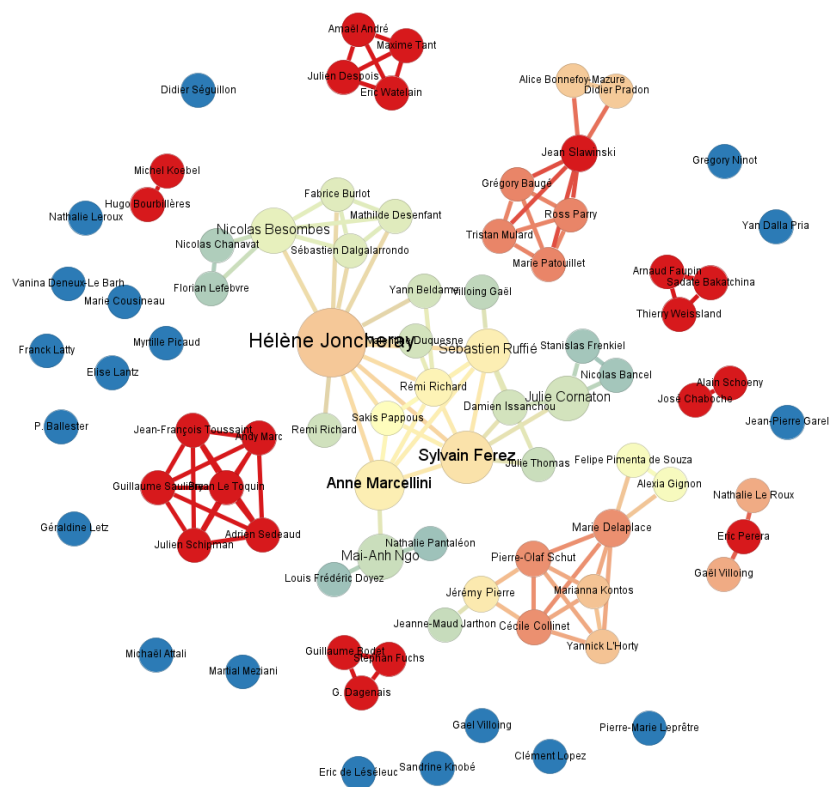


Figure 3 : Réseaux d'auteur de recherche sur le handisport

Répartition en France

Concernant la visualisation des lieux qui étudient le handisport en France, nous avons choisi de le faire à l'aide d'une carte de France (Figure 4).

Cette partie a soulevé plusieurs problèmes :

- les adresses d'universités sont peu représentatives (Lyon 1 à plusieurs localisations)
- les articles peuvent avoir 1 ou plusieurs d'adresses différentes -> doit-on en prendre une au hasard ou toutes ?

Le choix a donc été fait de se baser sur les codes postaux afin de ne pas avoir une adresse différente à chaque personne. Néanmoins, il y a toujours des mauvais résultats à cause de codes postaux étranger ou de boîtes postales formatées comme des codes postaux.

On s'aperçoit que Paris est la ville la plus active dans la production d'articles scientifiques sur le handisport, avec 118 participations à des articles en tout. Étonnamment, c'est la ville de Montpellier qui suit avec un assez grand nombre de participations également (40). Pour le reste, la répartition est assez homogène dans les autres villes de France.

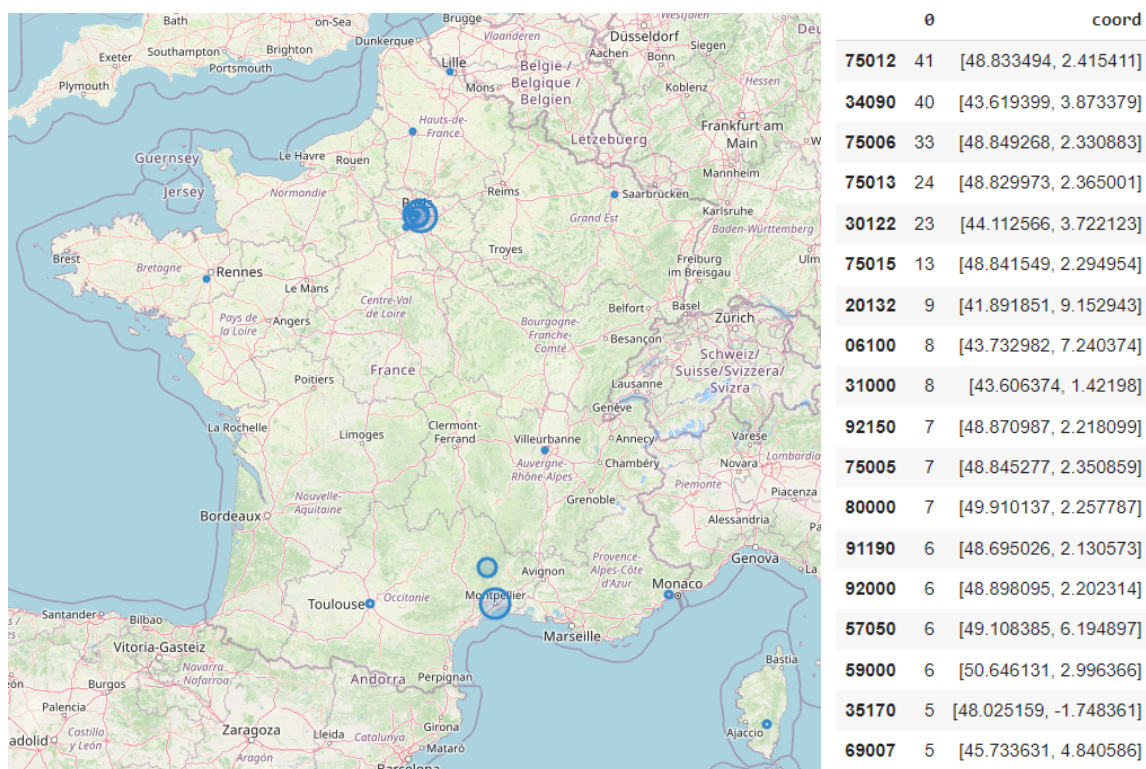


Figure 4 : Répartition géographique des chercheurs travaillant sur le handicap en France.

Type de handicap traité

Notre analyse s'est aussi portée sur la détermination du type d'handicap le plus fréquemment abordé au sein de l'archive HAL.

Ainsi, il est apparu que l'handicap mental était le plus récurrent dans cette archive, représentant 55,4% des occurrences recensées, tandis que l'handicap physique était répertorié à hauteur de 44,6%.

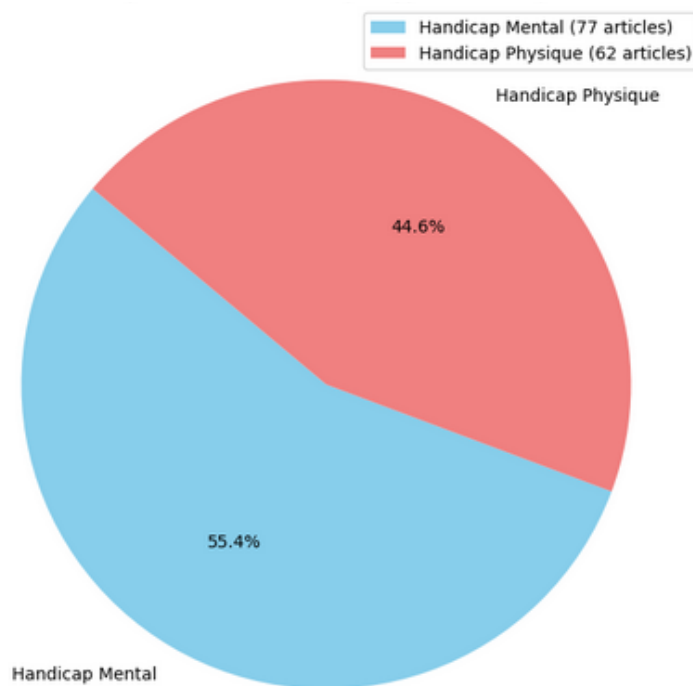


Figure 5 : Répartition des articles par type d'handicap

Pour obtenir ces informations, nous avons suivi une démarche en deux étapes.

Tout d'abord, nous avons utilisé un thésaurus afin d'identifier tous les termes employés pour décrire les handicaps physiques et mentaux.

Ensuite, dans un second temps, nous avons examiné la présence d'au moins un de ces termes dans les résumés des articles pour définir le type d'handicap abordé. (voir code en annexe 2)

Pour ce faire, nous avons utilisé la requête suivante :

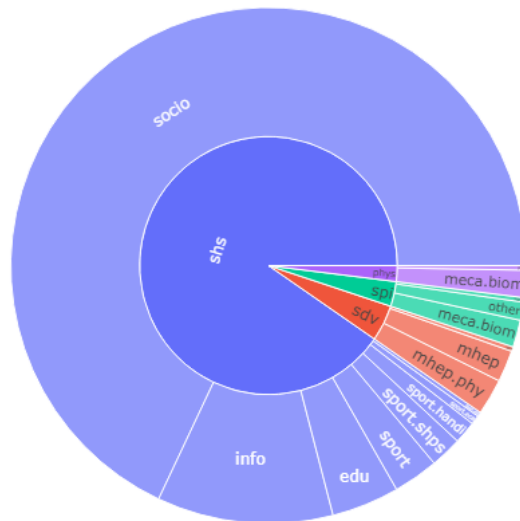
Cette méthodologie de recherche présente des limites, étant donné qu'un même article peut évoquer à la fois les deux types d'handicap dans son résumé, voire ne mentionner aucun des mots répertoriés dans le thésaurus. Néanmoins, elle offre l'avantage de formuler une hypothèse sur le type d'handicap le plus fréquemment abordé dans la littérature disponible dans HAL.

Conclusion et ouverture

Pour conclure sur ce rapport, on a pu voir que la recherche sur le handisport en France est assez récente et évolue assez rapidement depuis les années 2010. Les domaines d'études qui se penchent sur le sujet sont majoritairement les SHS. La répartition géographique des auteurs est plutôt homogène, avec tout de même une forte concentration à Paris et Montpellier. Ces aut.eurs.rices sont soit isolés, soit dans un petit groupe fortement lié, soit dans un grand groupe moyennement lié avec Hélène Joncheray au centre de ce groupe d'aut.eur.rice. Enfin, les types d'handicaps sont répartis équitablement dans les recherches.

Pour la suite des travaux, nous conseillons de partir de cette distinction des handicaps pour retravailler les études précédemment faites. De cette manière, regarder si l'évolution de cette équité est resté continu dans le temps ou dans l'espace.

Annexe



Annexe 1 : Représentation des domaines et spécialités de la recherche en handisport en fonction de leur occurrence.

```
# Exemple de catégorisation basée sur des mots-clés
as_csv['Type de Handicap'] = ''

# Catégorisation des handicaps mentaux
notnul = as_csv.dropna(subset="abstract_s").copy()
notnul.loc[notnul['abstract_s'].str.contains('mental|psycholog*|déficien
ce intellectuelle|QI|Trouble du spectre autistique|TDAH|TSA|',
case=False), 'Type de Handicap'] = 'Mental'

# Catégorisation des handicaps physiques
notnul.loc[notnul['abstract_s'].str.contains('phys*|moteur*|Kinésithéra
pie|Ergothérapie|Électrostimulation|Spasticité|Myopathie|Atrophie
musculaire|Orthèse|Prothèse|Amputation|Tétraplégie|Paraplégie|bles*',
case=False), 'Type de Handicap'] = 'physique'
```

Annexe 2 : code de la Figure 5

https://colab.research.google.com/drive/115Ct1Oe-rjLrw5QuR_ilPFVcRttvkKd1?usp=sharing

 Code Handisport.ipynb