Cartographie des sciences de l'alimentation au travers des publications académiques(1980-2018): vers quelles innovations les connaissances scientifiques orientent les légumineuses à graines en Europe ?

Tristan Salord, Marie-Benoît Magrini, Guillaume Cabanac, Marie-Josephe Amiot-Carlin, Marc Anton, Adeline Boire, Jean-Michel Chardigny, Matteo Lascialfari, Valérie Micard, Christophe Nguyen-Thé, et al.

Cartographie des sciences de l'alimentation au travers des publications académiques(1980-2018): vers quelles innovations les connaissances scientifiques orientent les légumineuses à graines en Europe ?

Tristan Salord[1], Marie-Benoît Magrini[1], Guillaume Cabanac[2], Marie-Josephe Amiot-Carlin[3], Marc Anton[4], Adeline Boire[4], Jean-Michel Chardigny[5], Matteo Lascialfari[1], Valérie Micard[6], Christophe Nguyen-Thé[7], Stéphane Walrand[8]

[1] AGIR, INRA, Université de Toulouse, 31326 Castanet-Tolosan, France – tristan.salord@inrae.fr, marie-benoit.magrini@inrae.fr, matteo.lascialfari@inrae.fr
[2] IRIT, Université de Toulouse, 31062 Toulouse, France – guillaume.cabanac@univ-tlse3.fr
[3] MOISA, CIRAD, CIHEAM-IAAM, INRA, Montpellier SupAgro, Université Montpellier, 34060 Montpellier, France– marie-josephe.amiot-carlin@inrae.fr
[4] BIA, INRA, 44000 Nantes, France – marc.anton@inrae.fr; adeline.boire@inrae.fr
[5] DPTI, INRA, 75338 Paris, France – jean-michel.chardigny@inra.fr
[6] IATE, INRA, Montpellier SupAgro, CIRAD, Université Montpellier, 34060 Montpellier, France – valerie.micard@supagro.fr
[7] SQPOV, INRA, 84140, Avignon, France – christophe.nguyen-the@inrae.fr
[8] INRA, UNH, Unité de Nutrition Humaine, CRNH Auvergne, Université Clermont Auvergne, 63001, Clermont-Ferrand, France – stephane.walrand@inrae.fr

Abstract

Pulses are essential for the sustainability transition of agrifood system. Their nutritional values allow animal-based protein intake reduction and contribute to healthy diets. Moreover, increasing their cultivation reduces fertilizers' uses and so, greenhouse gases emissions. But in high-income countries, and particularly in Europe, pulses consumption is very low. Previous works explained the mechanisms that led to a strong lock-in on pulses compared to the development of major crops, calling for stronger research and innovation to develop alternative food based on pulses. Food innovation remains strongly linked to the scientific knowledge and understanding how research output could sustain food innovation dynamics for sustainability is challenging. Mobilising scientometrics, STS Science and Technology Studies(STS) and Food Sciences and Technology(FST) scholars, the aim of this interdisciplinary study is to conduct a first map of the pulses FST. Describing what are the main issues tackled by FST on pulses will give to stakeholders a better understanding of the opportunities and expectations that could emerge in the development of a new food industry sector on pulses, as it will help academic research to identify specific field of interest to invest into. We considered 16 main pulses to build a dataset of more than 2,000 scholarly publications including at least one European author(articles, book, book chapters, and reviews) between 1980 and 2018, retrieved from the Web of Knowledge(Clarivate Analytics). Based on "natural language processing" clusterisation method we defined the main research themes structuring of the FST on pulses. We show that pulses have become a specific subfield of interest for FST in Europe, organized around of a great variety of subjects, and we focus on identifying the current gaps and opportunities to question science and innovation policies for pulses.

Keywords: bibliographic data; scientometrics; agrifood research; legumes;

Résumé

Les légumineuses sont essentielles pour la transition vers des régimes alimentaires sains et durables. Leur richesse en protéines permet de réduire la consommation de produits animaux; et leur culture permet de réduire l'usage des engrais azotés, et donc, les émissions de gaz à effet de serre. Pour autant, dans les pays à revenu élevé, et en particulier en Europe, la consommation de légumineuses reste très faible. S'appuyant sur la théorie des rendements croissants d'adoption, de précédents travaux ont révélé une situation de verrouillage technologique de ces cultures par rapport à d'autres espèces majeures comme les céréales. Cette situation appelle un déverrouillage pour lequel le développement des connaissances scientifiques est essentiel. Ces connaissances peuvent favoriser le développement de produits alimentaires innovants pour relancer leur consommation et enclencher un cercle vertueux de rendements croissants d'adoption. Cette communication dresse, pour la première fois, une cartographie des connaissances scientifiques sur ces légumineuses dans le champ des sciences de l'alimentation, grâce à un travail interdisciplinaire de chercheurs du domaine. A partir d'extractions du Web of Knowledge(Clarivate Analytics), un corpus de plus de 2000 publications scientifiques européennes sur 16 espèces de légumineuses entre 1980 et 2018 a été construit. L'analyse de ce corpus révèle la structure de ce champ scientifique et identifie les domaines-clés récemment développés, porteurs d'opportunités de marché. Nous observons une percée des enjeux de durabilité, qui forgent de nouvelles promesses sur lesquelles les innovations agroalimentaires sont susceptibles de se développer. Ce travail interroge plus largement les relations entre les politiques scientifiques et d'innovation pour les légumineuses afin de consolider cette émergence.

Mots clés : scientométrie, légumineuses à graines, agroalimentaire, lexicométrie

1. Introduction

La transition vers la durabilité des systèmes agroalimentaires repose sur un renouvellement des connaissances scientifiques, et ce, tout particulièrement dans le champ des Food Science and Technology(FST). Nos choix alimentaires orientent en effet la production agricole, et face aux enjeux écologiques, concevoir des aliments favorables à des systèmes agricoles qui préservent les ressources environnementales est essentiel (Frison 2016; Sabaté 2019; Tilman et Clark 2014). Le 21ème siècle tourné vers l'urgence écologique, sera peut-être celui d'une plus grande diversification alimentaire via le développement d'innovations alimentaires sur des espèces jusqu'ici peu consommées, mais dont les propriétés nutritionnelles et environnementales peuvent soutenir une plus grande durabilité. Parmi ces espèces, les légumineuses à graines sèches(dénommées sous la nomenclature de *pulses*[1]) font l'objet d'incitations croissantes des autorités publiques pour leur développement[2].

Riches en protéines, ces légumineuses permettent de rééquilibrer nos apports en protéines animales et végétales, en sus de leurs autres apports essentiels comme les fibres. Leur culture ne nécessite pas d'engrais azotés, et contribue de la sorte à réduire les gaz à effet de serre du secteur agricole(Peoples et al. 2019), soutenant ainsi une agriculture plus écologique. Leur consommation reste cependant très faible en Europe(environ 4 kg/an/hab) et face au verrouillage technologique dont ces espèces font l'objet, d'importantes innovations agroalimentaires sont nécessaires pour favoriser leur consommation (Marie-Benoit Magrini et al. 2018).

1 La dénomination pulses est une dénomination internationale n'ayant pas son équivalent direct en français. Les pulses rassemblent les légumineuses à graines récoltées en sec et dont l'usage n'est pas lié à l'extraction de l'huile, excluant le soja et l'arachide par exemple. En France(comme en Europe), la traduction de pulses n'existe pas directement car elle renvoie au rassemblement de deux classes d'appellation: les légumes secs traditionnellement consommés en alimentation humaine et les protéagineux dont l'usage majeur est l'alimentation animale. La nomenclature Pulses rassemblant les deux reflète une nomenclature unifiée plus tournée vers l'alimentation humaine et que nous conserverons dans cet article.

2 L'année internationale des pulses, organisée par les Nations Unies en 2016, en est une illustration marquante.

Renverser cette situation de verrouillage technologique reste difficile, mais un consensus de la littérature des STS(Sciences and Technology Studies) est que les connaissances scientifiques sont le ciment des innovations futures(Callon, Rip, et Law 1986; Dosi et Nelson 2010). Les sciences développent et renouvellent les connaissances permettant, au travers de leur combinaison, de construire de nouvelles innovations. Les approches évolutionnistes ont aussi particulièrement mis en avant comment la construction de promesses -« *expectations* »- (Arthur 2009) fondées sur de nouvelles connaissances est essentielle à l'émergence de nouvelles trajectoires technologiques. Comprendre la manière dont se construisent les sciences est donc essentiel pour éclairer les politiques de soutien à la recherche et l'innovation.

Dans cette perspective, de nouvelles activités de recherche en scientométrie se sont construites pour évaluer, mesurer et représenter les différentes sciences. Le développement de bases bibliographiques recensant les publications scientifiques(e.g, le Web of Science, SCOPUS), considérées comme un résultat mesurable des activités de recherche, couplé à celui des outils d'analyses statistiques des données textuelles, permet désormais de disposer d'outils performants pour représenter et analyser la construction de différents champs scientifiques (Glänzel et al. 2019). Une diversité de méthodes se sont développées au fil des dernières décennies, fondées principalement sur des analyses en termes de réseaux sémantiques et socio-sémantiques. Pour autant, dans le domaine scientifique nous intéressant, les FST, peu de travaux existent même si depuis une dizaine d'années quelques tentatives d'investigations scientométriques ont vu le jour (Acosta et al. 2017; Borsi et Schubert 2011). La présente étude s'inscrit donc dans la lignée de ces travaux, tout en s'efforçant d'apporter des éléments de réponses aux constats dressés précédemment.

En premier, l'enjeu est de déterminer comment les domaines de recherche en FST sur les pulses se structurent et ont évolué sur les dernières décennies, pour comprendre quels sont les nouveaux socles de connaissances qui peuvent soutenir le développement d'innovations agroalimentaires sur les pulses. Ensuite, l'enjeu est aussi de discuter, au regard de ces domaines de recherche investis sur les pulses, la manière dont ils constituent des leviers importants pour penser la durabilité future des régimes alimentaires. Troisièmement, au travers de l'exemple des pulses, ce travail développe une méthode originale permettant de décrire l'organisation d'un domaine scientifique de recherche. Ce travail ouvre la voie à l'identification de la manière dont les FST se structurent en sous-domaines de recherche, permettant d'éclairer les décideurs en charge de conduire les politiques scientifiques. Quatrièmement, ce travail ciblé sur les pulses a permis de préciser plusieurs thésaurus, l'un lié aux expressions scientifiques ou d'usage des espèces, l'autre lié aux termes-clés du champ des FST. Il n'existe pas de dictionnaire présentant une taxonomie internationalement reconnue de ces expressions. Aussi ce travail contribue à enrichir la construction d'un vocable précis et exhaustif, tant pour faire connaître la diversité des espèces de légumineuses, que pour décrire le champ des FST, pouvant être réemployé dans de futurs travaux.

L'interdisciplinarité est très présente dans cette étude qui a mobilisé les compétences d'experts des FST, des pulses, de traitement des données textuelles, de moissonnage et construction de bases bibliographiques. L'originalité de ce travail repose aussi sur la collaboration étroite entre chercheurs des STS et des FST. Ceci a permis, d'une part, de construire une base de données dans laquelle « le bruit » est fortement réduit; et d'autre part, de conduire une analyse concise de la construction du champ des pulses au sein des FST. Bien souvent les travaux en scientométrie sont conduits sans le couplage de cet ensemble de compétences, conduisant à des travaux imprécis(problème de délimitation du champ investi)

ou entachés d'erreurs de moissonnage ou d'interprétation. En contraste, notre travail interdisciplinaire a permis de construire un corpus d'un peu plus de 2 000 notices(articles, livres et chapitres d'ouvrages) traitant de domaines de recherche relatifs aux FST et aux pulses, et mobilisant au moins un auteur issu d'un pays européen. Ce corpus a été construit à partir d'une extraction d'un plus large corpus moissonné sur le WoS(Web of Science, Clarivate Analytics) contenant plus de 100 000 notices et couvrant l'échelle mondiale: nous renvoyons à l'article Magrini et al.(Marie-Benoît Magrini et al. 2019) présentant ce plus large corpus, la méthodologie mise en œuvre pour le construire et les raisons d'utilisation de la plateforme du WoS. Ces travaux (Marie-Benoît Magrini et al. 2019) visaient à mettre en relief le poids du soja dans les publications scientifiques(plus de 45% des citations pour cette espèce majeure au sein des légumineuses à graines) comparativement à l'ensemble des autres espèces légumineuses(pulses) cultivées dans les climats tempérés.

L'objectif de la présente communication est d'approfondir l'analyse du sous-corpus relatif aux FST sur les pulses et de proposer une méthodologie de construction d'une épistémologie des sciences à partir d'une clusterisation des sujets abordés dans les publications scientifiques.

2. Corpus et Méthode

2.1 Corpus de depart

De façon similaire aux travaux de Borsi et Schubert(Borsi et Schubert 2011), nous définissons les FST comme les sciences s'intéressant aux domaines des technologies de production et de transformation alimentaire, de la nutrition et de la consommation alimentaire. Fondé sur l'expertise de chercheurs issus des FST, le corpus de départ est le résultat de la fusion de quatre sous-corpus portant sur les thématiques de la transformation des aliments, de la nutrition humaine et ses effets santé, de l'acceptabilité par les consommateurs(perçue au travers de travaux d'analyses sensorielles) et des travaux liés au risque d'allergies.

Ces quatre sous-corpus[3] sont issus d'un précédent travail concernant la production scientifique mondiale traitant des légumineuses à graines auquel nous renvoyons pour une description détaillée de la méthode suivie (Marie-Benoît Magrini et al. 2019). Ils ont été constitué à partir du WoS et totalisent 8558 notices(articles, livres et chapitres d'ouvrages) publiées entre 1980 et 2018, pour un ensemble d'espèces de légumineuses à graines à l'échelle monde. Ces notices sont essentiellement en anglais[5] et viennent accompagnées d'un ensemble de *metadonnées*(Titre, Authors' keywords, Abstract, Authors' countries) qui ont été conservées pour la présente analyse[6].

3 Ces quatres corpus sont disponibles sur le portail des dataset INRAE : https://data.inra.fr/dataverse/root?q=pulses

5 Une limite de ce travail tient dans le fait que la plateforme bibliométrique du WoS tend à référencer préférentiellement les productions scientifiques anglophones, nous conduisant à ne pas intégrer les revues nationales écrites dans la langue d'usage de leur pays d'appartenance. Cependant, nous serions alors confrontés à un enjeu méthodologique de taille en termes de traitement automatisé de corpus multilinguistiques.

6 Il est important de noter que certaines metadonnées n'existent que depuis des périodes récentes(abtracts sont indexés dans le WoS depuis 1991).

2.2 Corpus retenu pour l'analyse

A partir de ce corpus de départ sur les FST, un nouveau corpus nettoyé et normalisé a été élaboré pour conduire nos analyses.Un premier filtre a consisté à ne garder que les notices portant sur les pulses(les notices portant sur le soja ainsi que sur les lathyrus-vicia ont été enlevées[7][8]).

L'aire géographique du corpus a également été restreinte à l'Europe. N'ont ainsi été gardées que les notices ayant au moins un auteur dont l'institution de rattachement appartient à un pays européen. Nous partons du consensus de la « géographie de l'innovation » selon lequel la diffusion des connaissances reste très liée à la proximité géographique des individus les possédant. Partant, se pencher sur l'analyse des connaissances scientifiques pour interroger la capacité de l'Europe à développer un nouveau secteur agroalimentaire des pulses, nous à amener à restreindre le corpus à cette aire géographique(Figure 1 - « Filtrage espèces et pays »). A la suite à ce premier travail de filtrage, le corpus s'est réduit à un ensemble de 2 677 notices.

Afin d'accroître la robustesse du corpus pour l'analyse, une étape intermédiaire de nettoyage et de normalisation a été conduite pour repérer, d'une part, les termes à exclure des futures analyses lexicométriques[9], et, d'autre part, les termes devant être normalisés ou désambigüisés[10]. Cette opération a été rendue possible grâce à un algorithme de normalisation travaillant à partir d'un dictionnaire de règles co-construites avec les experts(plus de 340 règles ont été définies), qui a été augmenté de façon incrémentale par les différentes passes de l'algorithme et les résultats d'analyses préliminaires conduites avec le logiciel Iramuteq.

Plusieurs « analyses des similitudes[11] » ont été en effet conduites sur le corpus de façon à repérer, dans les graphes de « co-occurrences de termes », les termes problématiques, les mots-outils, les expressions composées(qui auraient alors été traitées de façon séparées et non unitaire par le logiciel)(Figure 1 - Processus de normalisation du corpus, "Analyse des similitudes"), afin d'affiner les analyses.

Cette séquence d'opérations « normalisation / analyses / mise à jour du dictionnaire de règles » a été répétée trois fois, jusqu'à obtention de résultats jugés satisfaisants par les experts.

Enfin, une dernière étape de nettoyage a eu lieu, sur la base d'une première Classification Hiérarchique Descendante(CHD, expliquée ci-après) du corpus(voir Figure 1 - « Classification Hiérarchique Descendante »). La CHD nous a permis de vérifier la pertinence des revues et articles composant les classes[12], et de repérer parmi les notices ne rentrant pas dans le scope du champ analysé(cf. section 2.1). C'est sur la base de ce dernier nettoyage que le corpus a été finalisé pour conduire la clusterisation finale représentant le champ des FST sur les pulses et commentée dans la section 3.

7 Ces dernières étant non présentes dans les usages européens, il a été décidé de les exclure du corpus.

8 Pour plus d'explicitation sur le filtrage par espèce voir Magrini et al. (Marie-Benoît Magrini et al. 2019)

9 Par exemple des unités de mesure, des expressions sous forme d'équation, des expressions trop génériques telles que « this study presents an analysis of »

10 Par exemple, des expressions synonymes entre elles, notamment liées aux écritures avec sigle ou sans sigle, les orthographes variables de même termes, les problèmes d'homonymes, etc.)

11 Pour la description de la méthode de l'analyse des similitudes voir (Ratinaud et Marchand 2011)

12 Nous utilisons de façon synonyme les termes de classe ou de cluster, de classification ou de clusterisation.

Le corpus final représentant les FST sur les pulses impliquant l'Europe comprend 2386 notices, soit 28% des notices du corpus de départ qui portaient sur un ensemble de légumineuses à graines(dont soja) à l'échelle monde.

2.3 « Clusterisation » des notices

Il existe de nombreuses méthodes de traitement automatisé du langage pour analyser le contenu sémantique d'un champ scientifique donné, et au sein de ces dernières on trouve autant de techniques et algorithmes de « clusterisation » permettant de découper dans ces univers sémantiques des sous-ensembles(decision tree, pattern(rule)-based classifiers, SVM classifiers, bayesian classifiers, etc.).

Deux raisons principales nous ont poussé à choisir ici la méthode Reinert mise en œuvre dans le logiciel Iramuteq :(a)il s'agit d'une méthode "déterministe"[13] de clustering, assurant par là-même la reproductibilité de nos résultats,(b) cette méthode prend en compte les contextes de citation des termes(appelé u.c.e[14]).

La méthode Reinert (Max 1993; Reinert 1990) est une méthode de clusterisation de textes(ou CHD pour Classification hiérarchique descendante) qui consiste à confronter l'occurrence de termes(dénommés « formes ») au regard de leur contexte d'énonciation(ie. de « segments de texte », ST[15]). Une des spécificités de cette méthode d'analyse est qu'elle prend également en compte la fonction grammaticale[16] des termes. Ainsi sont exclus des analyses tous les mots-outils(tous les déictiques, déterminants, etc.), celles-ci se basant uniquement sur les termes/formes dites « pleines »(verbes, noms, adjectifs, adverbes)[17]. A la fin du processus de clusterisation, les analyses se présentent sous la forme d'une arborescence de classes reprenant les formes actives les plus fréquemment associées et les plus représentatives de chaque classe (Ratinaud et al. 2019).

Les paramètres donnés en entrée à l'algorithme de Reinert sont les suivants :

- nombre de classes demandées en fin de phase 1 : 25
- Nombre maximum de formes analysées : 10 000

Ceci étant, il est important de noter que les résultats présentés dans les sections suivantes(sections 3.2 à 3.4), ont été stabilisés d'une part, grâce au travail d'interprétation menés par les experts à chaque fois qu'un résultat intermédiaire était produit, et d'autre part, à l'aide d'une série de tests nous ayant permis de noter quelles classes étaient les plus « robustes » aux paramètres de clusterisation donnés en entrée au logiciel(i.e, les classes les plus stables quels que soient les paramètres de clusterisation appliqués).

13 L'algorithme lancé plusieurs fois avec les mêmes paramètres de configuration reproduira les mêmes résultats.

14 « u.c.e » signifiant « unité de contexte élémentaire ». Se reporter notamment à (Ratinaud et Marchand 2012)

15 Un segment de texte correspond à une chaîne de plus ou moins 40 caractères se terminant par un signe de ponctuation conclusif. La taille de la chaîne de caractère, l'ampleur du contexte d'énonciation, est paramétrable. Le choix des 40 caractères repose sur un constat statistique.

16 Ou « POS » pour « part of speech »(tagging), terme technique utilisé dans le domaine du traitement automatisé du langage pour qualifier la fonction d'une « forme », on peut également parler d'étiquetage morpho-syntaxique.

17 Pour ce faire le logiciel se base sur un ensemble de dictionnaires donnant, notamment, pour chaque terme sa fonction grammaticale. Nous avons utilisé dans le cadre de nos analyses une version modifiée du dictionnaire anglais du logiciel afin d'augmenter le filtrage de certains termes(notamment toutes les unités de mesure anglaise ainsi que les termes rhétoriques de l'écriture scientifique).

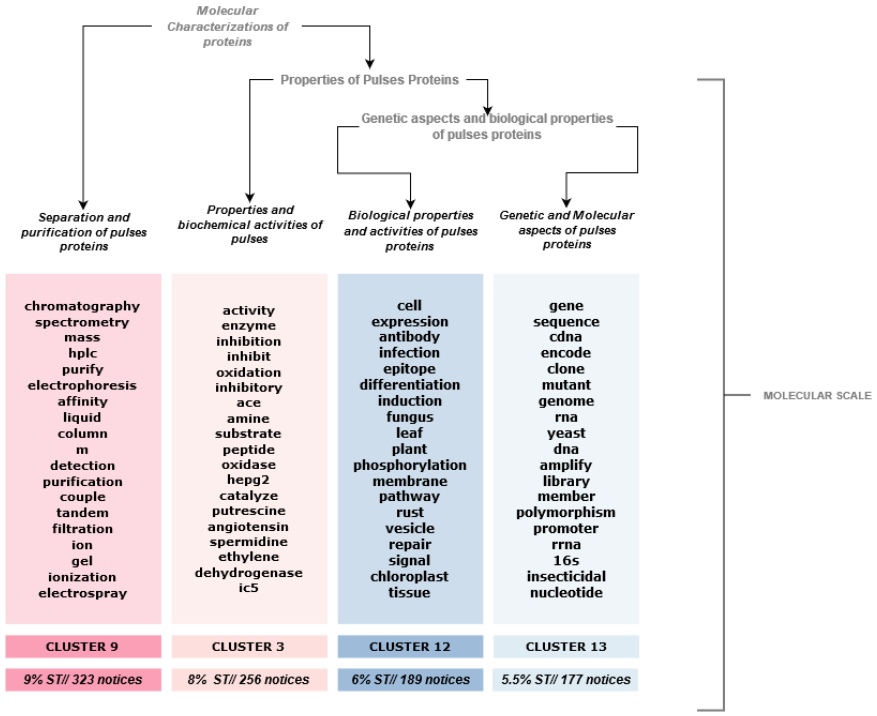
3. Résultats et discussion

3.1 Structuration des champs thématiques des FST sur les pulses

La figure suivante(Figure 2) intitulée « *Dénomination des classes et ensembles thématiques du dendrogramme(corpus final)* » représente le résultat de la CHD appliquée au corpus final. Cette clusterisation repose sur l'identification par Iramuteq de 18 783 formes(expressions retenues par le logiciel en fonction des règles définies) qui se répartissent dans 10 371 segments de texte représentant 87% de l'ensemble des segments de texte du corpus. L'ensemble de ces 18 783 formes totalisent 366 158 occurrences d'apparition.

La CHD retenue conduit à répartir ces formes[18] en 13 classes. Pour améliorer l'interprétation des résultats, les couleurs de chaque classe ont été changées par un dégradé de tons[19] indiquant leur taille relativement au pourcentage de segments de texte participant à la classe.

L'interprétation de cette classification a fait l'objet de plusieurs réunions d'analyse entre experts afin de donner un sens au découpage du corpus FST en classes thématiques. Il s'est agi, de façon ascendante : i) de qualifier l'**« unité thématique »** de chaque classe ; ii) puis de nommer chacun des nœuds supérieurs, conduisant à identifier également des « **ensembles thématiques** » plus vastes. Pour ce faire, le travail des experts s'est fondé sur l'extraction des 15 notices les plus sur-représentées dans chacune des 13 classes. Les dénominations choisies sont reportées sur le schéma et commentées ci-après.



18 Les formes apparaissant dans les classes sont celles dont l'occurrence dans l'ensemble des segments de texte est supérieur à 3 et dont la p value dans la classe est supérieure à 0.5. Elles sont par ailleurs classées par ordre décroissant(en fonction de leur chi2) dans chacun des classes.

19 Allant du bleue pour les valeurs des premiers quartiles, au rouge pour les valeurs des derniers quartiles.

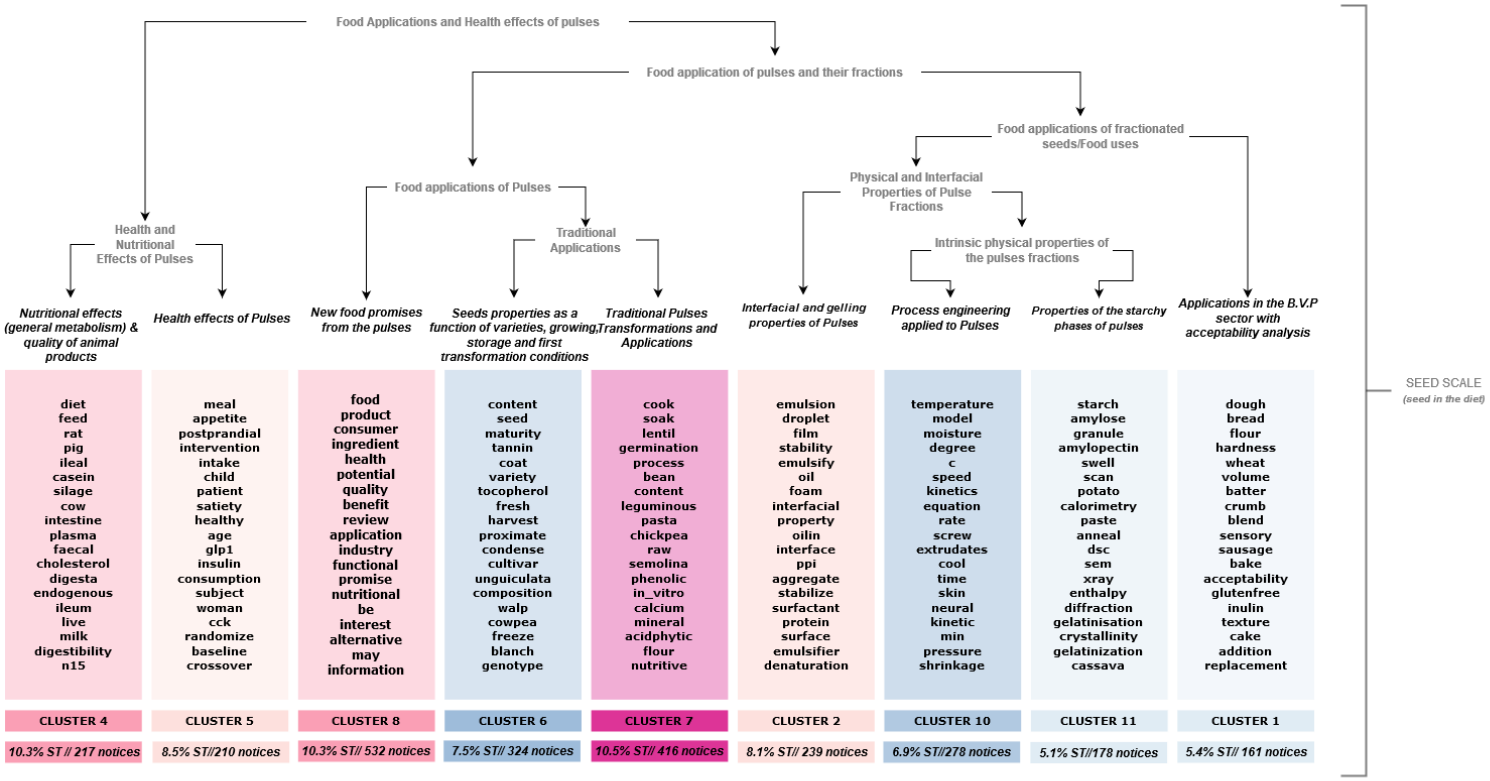


Figure 1: Dénomination des classes et ensembles thématiques (corpus final)

Le corpus bibliographique étudié présente deux grands ensembles thématiques relatifs à l'échelle des travaux de recherche en FST. Un 1er ensemble thématique(schéma haut de la Figure 2) porte sur la « ***caractérisation moléculaire des protéines de pulses*** »(28,5% des segments de texte du corpus ; classes 9,3,12,13). Un second(schéma bas de la figure 2), traite plus des propriétés des graines de pulses liées à leurs « ***applications alimentaires et effet santé*** »(72,6% des segments de texte du corpus ; classes 4,5,8,6,7,2,10,11 et 1). Cette séparation thématique se révèle également dans l'analyse temporelle(cf. section 3.3). Notons que ces deux grands ensembles de travail renvoient également à des méthodes/des techniques d'investigation scientifique différentes, mobilisant par exemple directement le facteur humain dans le protocole de recherche ou pas.

Le premier ensemble thématique sur la caractérisation moléculaire des pulses, se subdivise lui-même en deux champs scientifiques: l'un s'intéresse plus aux ***méthodes de séparation et de purification*** des protéines de pulses(cluster 9 – 9% des segments de texte), tandis que l'autre porte plus spécifiquement sur la ***caractérisation des propriétés des protéines de pulses(***clusters 3,12,13 – 19,5% des segments de texte). Ce second ensemble distingue à son tour deux sous-ensemble: l'un centré sur les « ***propriétés et activités biochimiques des***

pulses »(cluster 3 – 8%des segments de texte) ; l'autre sur les « ***aspects génétique et propriétés biologiques des protéines de pulses*** »(clusters 12 et 13).

A la lecture des résumés d'articles scientifiques, le second ensemble thématique(sur les applications alimentaires et effets santé) nous paraît mis en tension par deux types de focale selon que les productions scientifiques s'attachent :

- à ***étudier/décrire les effets santé de la consommation des pulses***(clusters 4 & 5 – « Effets santé et nutritionnels des pulses », 18,8% des segments de texte),
- à identifier les apports potentiels des pulses aux enjeux et problématiques alimentaires contemporaines– « ***Applications alimentaires des pulses et de leurs fractions*** »(clusters 8,6,7,2,10,11 et 1, soit 53,8% des segments de texte).

Plus précisément, les classes 4 et 5 révèlent fortement les préoccupations de santé publique des pays occidentaux. La classe 4 renvoie ainsi plus à des problématiques de santé par l'alimentation(ou alimentation saine), tandis que la classe 5 se focalise plus sur les effets santé des pulses dans le cadre des maladies chroniques. Notons ici que la lecture des résumés des articles(contribuant le plus aux classes) reste essentielle pour l'interprétation de celles-ci.

Le deuxième type de focale(« ***Applications alimentaires des pulses et de leurs fractions*** ») apparaît quant à lui plus contrasté dans les univers sémantiques qu'il mobilise. En effet, deux logiques différentes émergent selon que l'intérêt des études se porte plus : sur une consommation de légumineuses issue de pratiques alimentaires plus ***traditionnelles*** – comme la consommation de légumes secs-(cluster 6, 7 et 8 dans une moindre mesure, soit 28,3% des segments de texte), ou sur les atouts de leur utilisation par l'***industrie agro-alimentaire***(clusters 10, 11, 1, soit 25,5% des segments de texte). A noter que les clusters 10 et 11 renvoient en particulier aux propriétés liées aux fractions de légumineuses que l'on ne retrouve pas dans des usages plus traditionnels.

Cette répartition entre pratiques plus traditionnelles(voir ordinaires) et pratiques industrielles se voit d'autant mieux si l'on fait contraster les clusters 7 et 1, le premier rassemblant les travaux menés autour des usages et pratiques de transformation traditionnelles des pulses, tandis que le second se centre sur les avantages texturaux et/ou sensoriels que l'on peut tirer de l'usage des pulses dans le champ des B.V.P.

In fine, ce sous-ensemble thématique plus centré sur des problématiques agroalimentaires industrielles(clusters 2,10,11 et 1) met en relief différentes pistes d'exploitation des pulses. Des travaux s'intéressent ainsi aux pulses pour leur « ***propriétés interfaciales et gélifiantes*** »[20](cluster 2 et 8, soit1% des segments de texte), comme une source spécifique d'amidon(cluster 11, « ***propriétés des phases amylacées des pulses*** », 5,1%), pour les avantages que l'on peut en tirer en termes de propriétés texturales - et notamment rhéologique-, sensorielles des aliments(cluster 1, « ***Applications du secteur B.V.P avec études d'acceptabilité*** » 5,4%). Bien sûr, l'extraction de ces propriétés implique la mise en œuvre de techniques ad hoc d'extraction, de fragmentation, de purification, etc. que l'on peut retrouver dans le cluster 10, que nous avons, à ce titre, intitulé « ***Génie des procédés appliqués aux pulses*** »(6,9% des segments de texte du corpus).

20 Par exemple, la notice la plus représentative de la classe 2(WOS :000314374700009), s'intéresse aux effets – et à l'amélioration- des films de protéines de pois par adjonction d'un éther de cellulose spécifique: « The effect of the addition of carboxymethylcellulose(in three viscosity types: CMC 30, CMC 1000, CMC 10000), hydroxypropylmethylcellulose(HPMC), pectin(PEK), and chitosan lactate(MCH) on the properties of pea protein films was investigated. »

3.2 Identification des Fronts de Science à partir de l'analyse temporelle

La projection des dates de publication de notices les plus sur-représentées dans chaque classe(ou *chronodendrogrammes*, voir les figures 3 et 4 suivantes) nous permet de nous rendre compte de l'évolution temporelle(par année) des univers thématiques que nous venons d'identifier(section précédente) et de repérer, par là même, ce que nous avons appelé ici des *fronts de science*[21]. L'interprétation conjointe de ces deux dendrogrammes temporels nous permet d'apprécier l'évolution des champs investis par les FST sur les pulses.

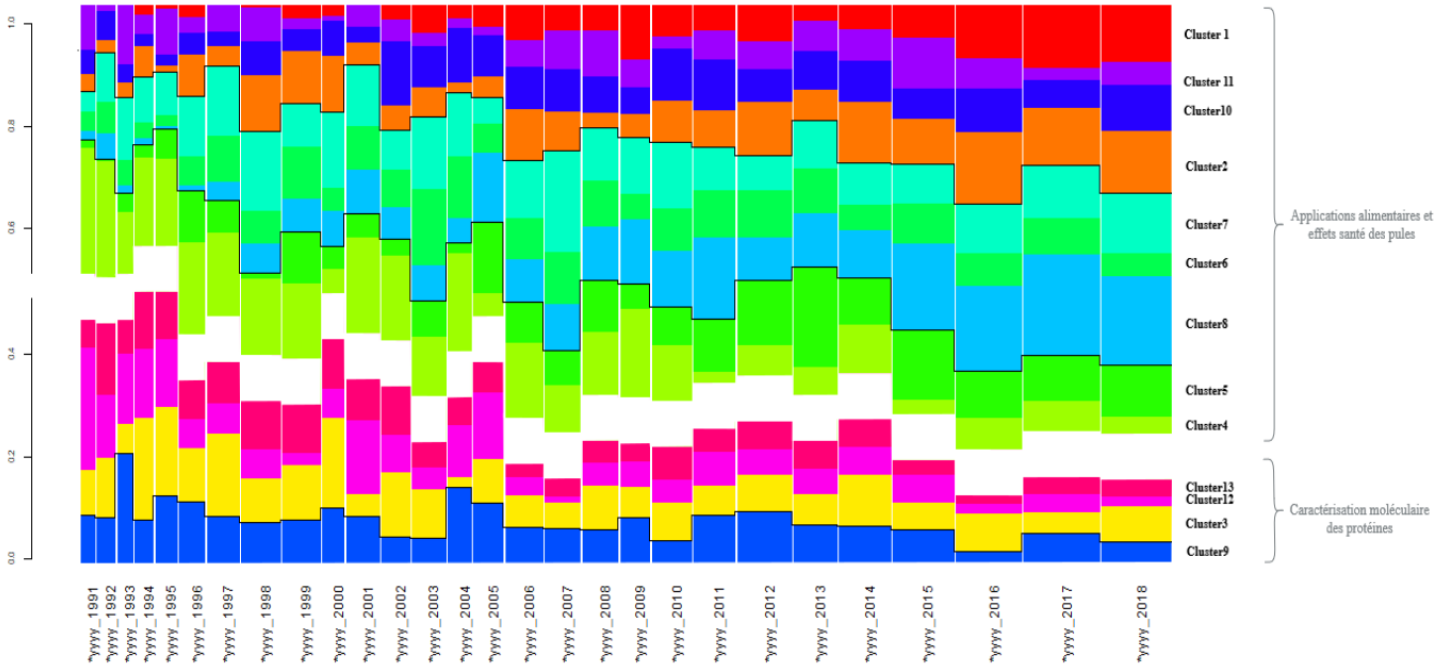


Figure 2: Chronodendogramme des classes(par proportion) - vue éclatée

Ce chronodendrogramme représente l'évolution annuelle de chaque classe. La taille de chaque polygone est proportionnelle au nombre de segments de texte rassemblés dans une classe donnée, pour une année donnée. Nous avons pris soin ici de reproduire la bi-partition commentée précédemment(voir section 3.1)entre deux échelles de travaux (au niveau moléculaire, au niveau des graines de légumineuses) ; elle est représentée par l'espace blanc scindant le chronodendrogramme en deux ensembles.

On remarque ainsi à la lecture de cette figure, que les classes thématiques 1,2,8 et 5 concernant respectivement « *les applications des pulses dans le champ des B.V.P* », « *les propriétés interfaciales et gélifiantes des pulses* », « *les nouvelles promesses alimentaires des pulses* », et dans une moindre mesure, « *les effets santé des pulses* », sont particulièrement investies sur les dernières années. Les travaux les plus récents semblent donc s'intéresser majoritairement aux effets « bénéfiques » des pulses dans une alimentation aussi bien issue

21 Dans ce cas, la sur-représentation d'une classe sur plusieurs années constitue un indice de ce qui contitue à un instant donné un front de sciences.

des secteurs de l'agro-industrie que d'un cadre de formes de consommation plus traditionnelle, révélant ainsi la co-existence d'une recherche fondée sur des modèles alimentaires différents. Ces effets bénéfiques sont tant d'ordre *technologique*(utiliser des farines de pulses pour améliorer les qualités rhéologiques d'une pâte, par exemple), *gustatif(*rajouter des effets de texture ou de goût à des préparations alimentaires) ou de *santé*(par exemple, pour la régulation du diabète). On note également que la grande majorité des études d'acceptabilité par le consommateur se situent dans ces clusters sur-représentés dans les années les plus contemporaines du corpus.

Le poids contemporain de ces clusters souligne donc, à notre sens, une plus grande reconnaissance contemporaine des intérêts nutritionnels et technologiques des pulses pour l'alimentation future. Ceci peut être vu comme un stade d'avancement significatif de la recherche sur les pulses permettant de renvoyer un ensemble de promesses, dont l'industrie peut se servir pour développer des innovations alimentaires, dans un contexte sociétal très marqué par les préoccupations de santé publique par l'alimentation. En effet, par exemple, dans plusieurs groupes de travail de l'ANSES(agence européenne de la sécurité alimentaire) de ces dernières années, la question de l'intérêt des légumineuses pour l'alimentation a fait l'objet de débats plus fréquents, permis par l'accroissement des publications scientifique sur ce sujet. En particulier, nous avons vu que, sur les dernières années, la classe 8 "*des promesses*" est à la fois, celle la plus sur-représentée et celle regroupant le plus de notices.

Au-delà de ces classes sur-représentées(qualifiées de fronts de science), il reste intéressant de mettre en avant des classes qui restent très investies par la recherche. On peut ainsi observer(voir figure ci-après) que le cluster 7, significativement plus présent au début des années 2000, reste d'un intérêt important pour la recherche comparativement aux autres classes. Ce cluster portant sur des solutions traditionnelles de transformation/consommation des pulses[22] et traitant d'applications vers des produits de grande consommation courante(comme le développement de pâtes alimentaires à base de pulses) se trouve donc contrebalancé, aujourd'hui, par les classes *fronts de science* de travaux mettant en avant de nouvelles propriétés et promesses(health, alternative, glutenfree, etc.) des pulses plus utilisés au travers d'ingrédients insérés dans un spectre plus élargi d'aliments(Classes 2, 8 et 1 notamment pour les applications élargies en BVP dont les travaux ont significativement augmenté sur les années les plus récentes).

Par ailleurs, nous n'observons pas de progression des travaux des classes regroupées dans la thématique "caractérisation moléculaire des protéines" alors que ces classes étaient significativement sur-représentées jusqu'au début des années 2000(cf. Figure 4). Le poids des notices contribuant à ces classes reste en effet relativement stable au fil du temps, voir diminue légèrement comme on peut s'en apercevoir dans la figure suivante.

Cette figure rend compte du poids relatif de chaque classe en fonction du nombre de notices y contribuant pour une année donnée. Elle a été réalisé en rapportant chaque segments de texte classé dans une classe donnée pour une année donnée à la notice dont il a été extrait.

22 On y retrouve des travaux s'intéressant aux modes de trempage des légumineuses, de germination des graines,… On notera à ce propos que parmi les nouvelles promesses évoqués dans le corps de notre analyse, on observe une progression des formes liées aux questions de germination des graines, particulièrement présentes dans la classe 7, mais qui se retrouve aussi dans la classe 8(positionnement de la forme au-delà des 19 premières formes caractérisant chaque classe(cf Figure 1).

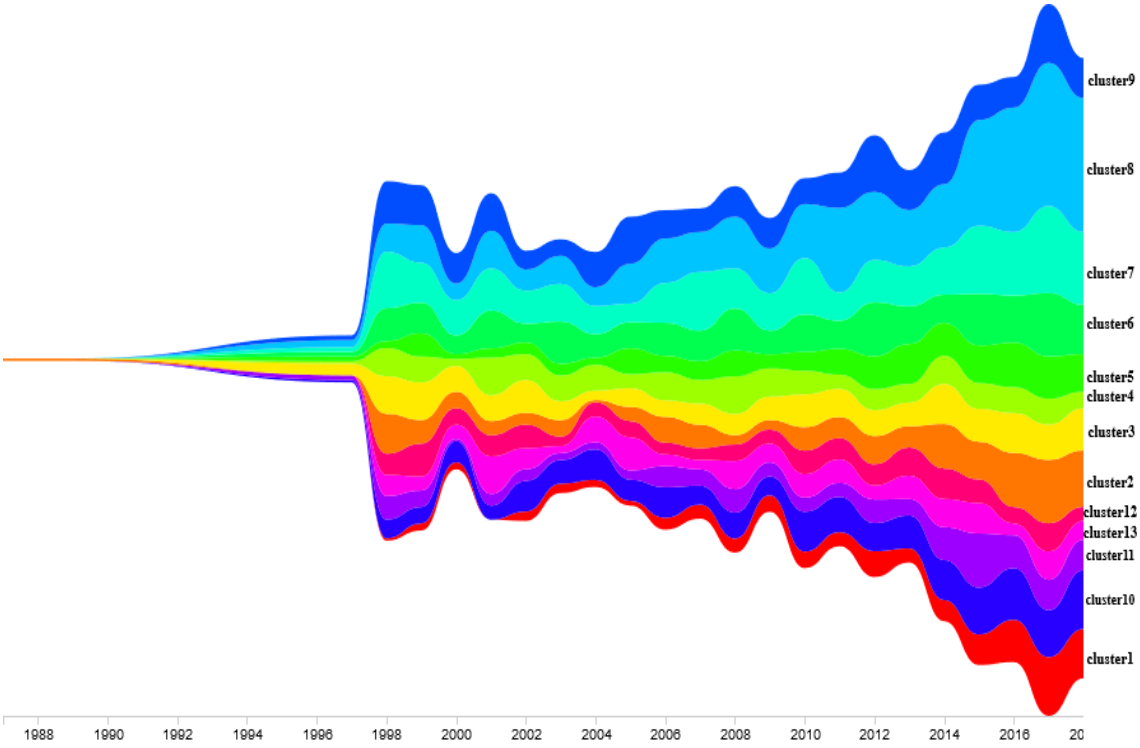


Figure 3 - Décompte des notices participant aux différentes classes

La séparation entre deux grands ensembles thématiques recouvrant, d'un côté, les travaux s'intéressant aux propriétés des molécules de protéines de pulses, et de l'autre côté, les effets des pulses dans le champ de l'alimentation et de la production agro-alimentaire, telle que nous l'évoquions dans la section précédente(section 3.1) semble donc se doubler d'une séparation temporelle. Une explication possible tient au fait que les progrès de connaissances dans la caractérisation moléculaire des protéines ont permis d'orienter les travaux de recherche vers des applications alimentaires concrètes et la reconnaissance des intérêts/promesses pour le consommateur.

Conclusion

L'analyse des mondes lexicaux des publications européennes portant sur les pulses dans le champ des FST nous a permis de mettre en avant la manière dont se structurent cet espace de recherche, ouvrant sur plusieurs perspectives.

D'abord, nous avons montré que deux grands corps correspondant à deux grandes échelle de recherche organisent cet espace scientifique, l'un s'intéressant plutôt à la caractérisation moléculaire des protéines de pulses, tandis que l'autre se penche plus sur les modalités d'inclusion et d'augmentation pulses dans l'alimentation, en situant une grande partie ces recherches à des enjeux de santé publique et de durabilité future.

Ce résultat ouvre des perspectives de recherche importante. Il serait intéressant de comparer ces résultats pour d'autres espèces afin d'apprécier si le positionnement décrit plus haut est propre aux pulses ou se généralise pour d'autres familles d'espèces. En particulier, cela permettrait d'évaluer si ce changement de paradigme des recherches en FST vers une « alimentation-santé » s'opère dans tous les groupes alimentaires ou si les pulses apparaissent comme les porteurs, voir les précurseurs de ce tournant.

Ensuite, la projection temporelle du développement de ce domaine scientifique (les pulses au sein des FST) nous a amené vers la construction d'une certaine historiographie du champ. Nous avons ainsi observé une orientation de la recherche d'abord plus soutenue sur la caractérisation moléculaire, avant de basculer vers des investissements plus forts sur les applications alimentaires, puis, plus récemment, sur la reconnaissance d'un ensemble de promesses technologiques et nutritionnelles pour le consommateur qui questionnent aujourd'hui leur acceptabilité.

La reconnaissance de nouvelles promesses autour des pulses (propriétés nutritionnelles et technologiques pour développer une alimentation-santé) par les institutions et les industriels est une condition essentielle d'engagement d'une nouvelle trajectoire industrielle dans ce secteur. Comme expliqué dans Magrini et al., 2019 le développement des promesses - expectations - est la condition première à l'émergence de rendements croissants d'adoption dans les processus des transitions, et la légitimation de ces promesses est d'autant plus forte que des travaux scientifiques en attestent.

Cette remarque appelle à mieux analyser la dimension géographique des liens entre progrès des connaissances scientifiques et innovations, en analysant la progression des innovations en termes de nouveaux produits agro-alimentaires sur le marché européen ; voir également en analysant les liens entre brevets et publications scientifiques citées.

Remerciements

Ce travail a bénéficié d'un financement du projet de recherche H2020 No 727672 LEGVALUE(Fostering sustainable legume-based farming systems and agri-feed and food chains in the EU) 2017-2021. Les auteurs remercient également les experts, Gaëelle Arvesinet, Colette Larré, ayant contribués à la construction des requêtes de recherché.

Références

Acosta, Manuel et al. (2017). « The Geography of University Scientific Production in Europe: An Exploration in the Field of Food Science and Technology ». Scientometrics 112(1): 215-40.

Allahyari, Mehdi et al. (2017). « A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques ». arXiv:1707.02919 [cs]. http://arxiv.org/abs/1707.02919 (22 janvier 2020).

Amiot, Marie-Josephe et al. « Nutrition and Grain-Legumes WoS DataSet 1980-2018 ». https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/5MI04S (27 janvier 2020).

Anton, Marc et al. « Processing and Grain-Legumes WoS DataSet 1980-2018 ». https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/VP7PRI (27 janvier 2020).

Arthur, W. Brian. (2009). The Nature of Technology: What It Is and How It Evolves. Publ. in Penguin Books. London: Penguin Books.

Arvisenet, Gaelle, Marie-Benoit Magrini, Hugues Leiser, et Guillaume Cabanac. « Acceptability and Grain-Legumes WoS DataSet 1980-2018 ». https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/PDXRYM (27 janvier 2020).

Borsi, Balázs, et András Schubert. (2011). « Agrifood Research in Europe: A Global Perspective ». Scientometrics 86(1): 133-54.

Callon, Michel, Arie Rip, et John Law. (1986). Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World. New York; Secaucus: Palgrave Macmillan Springer [distributor. https://link.springer.com/openurl?genre=book&isbn=978-1-349-07410-5 (24 janvier 2020).

Dosi, Giovanni, et Richard R. Nelson. (2010). « Technical Change and Industrial Dynamics as Evolutionary Processes ». In Handbook of the Economics of Innovation, Elsevier, 51-127. https://linkinghub.elsevier.com/retrieve/pii/S0169721810010038 (24 janvier 2020).

Frison, Emile A. (2016). From Uniformity to Diversity: A paradigm shift from industrial agriculture to diversified agroecological systems. IPES FOOD. http://www.ipes-food.org/_img/upload/files/Uniformiteala%20Diversite_IPES_FR_Full_web.pdf (24 janvier 2020).

Glänzel, Wolfgang, Henk F. Moed, Ulrich Schmoch, et Mike Thelwall. (2019). Springer Handbook of Science and Technology Indicators. Cham: Springer International Publishing. http://link.springer.com/10.1007/978-3-030-02511-3 (24 janvier 2020).

Larré, Colette, Sandrine Denery, Hugues Leiser, et Guillaume Cabanac. « Allergy and Grain-Legumes WoS DataSet 1980-2018 ». https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/BZG0R7 (27 janvier 2020).

Magrini, Marie-Benoit et al. (2018). « Pulses for Sustainability: Breaking Agriculture and Food Sectors Out of Lock-In ». Frontiers in Sustainable Food Systems 2: 64.

Magrini, Marie-Benoît et al. (2019). « Peer-Reviewed Literature on Grain Legume Species in the WoS (1980–2018): A Comparative Analysis of Soybean and Pulses ». Sustainability 11(23): 6833.

Max, Reinert. (1993). « Les “mondes lexicaux” et leur 'logique" à travers l’analyse statistique d’un corpus de récits de cauchemars ». Langage & société 66(1): 5-39.

Peoples, Mark B. et al. (2019). « The Contributions of Legumes to Reducing the Environmental Risk of Agricultural Production ». In Agroecosystem Diversity, Elsevier, 123-43. https://linkinghub.elsevier.com/retrieve/pii/B978012811050800008X (24 janvier 2020).

Ratinaud, Pierre et al. (2019). « Structuration des discours au sein de Twitter durant l’élection présidentielle française de 2017: Entre agenda politique et représentations sociales ». Réseaux n°214-215(2): 171.

Ratinaud, Pierre, et Pascal Marchand. (2011). « L’analyse de similitude appliquée aux corpus textuels : les primaires socialistes ».

Reinert, Max. (1990). « Alceste Une Méthodologie d’analyse Des Données Textuelles et Une Application: Aurelia De Gerard De Nerval ». Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique 26(1): 24-54.

Sabaté, Joan. (2019). Environmental Nutrition: Connecting Health and Nutrition with Environmentally Sustainable Diets.

Tilman, David, et Michael Clark. (2014). « Global Diets Link Environmental Sustainability and Human Health ». Nature 515(7528): 518-22.