



HAL
open science

Recherches doctorales intéressant la formation des adultes : noyau et cercle élargi de thèses soutenues aux États-Unis de 2000 à 2022, comparaison avec la France

Olivier Las Vergnas, Patrick Bury, Christophe Jeunesse

► To cite this version:

Olivier Las Vergnas, Patrick Bury, Christophe Jeunesse. Recherches doctorales intéressant la formation des adultes : noyau et cercle élargi de thèses soutenues aux États-Unis de 2000 à 2022, comparaison avec la France. *Savoirs : Revue internationale de recherches en éducation et formation des adultes*, 2023, Vie de la recherche en formation des adultes, N° 60 (3), pp.95-133. 10.3917/savo.060.0095 . hal-03988340

HAL Id: hal-03988340

<https://hal.science/hal-03988340>

Submitted on 14 Feb 2023

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.

Recherches doctorales intéressant la formation des adultes :
Noyau et cercle élargi de thèses soutenues aux USA de 2000 à 2022, comparaison avec la France.

Olivier LAS-VERGNAS (1,2), Patrick BURY (3), Christophe JEUNESSE (1)

1 : CREF, équipe ApForD, Université Paris-Nanterre

2 : CIREL, équipe Trigone, Université de Lille

3 : Société Clever Mind

Résumé : Cet article explore les thèses de doctorat de recherches soutenues aux USA entre 2000 et 2022, concernant la formation des adultes et accessibles dans la base PQDT. Il complète une première étude publiée dans le précédent numéro de la revue *Savoirs*. Il s'appuie sur des analyses lexicales et bibliométriques pour proposer et étudier deux corpus de tailles différentes. Le premier (noyau = 1649 thèses) regroupe les thèses faisant explicitement référence au thème dans leur titre ou mots clefs, et le second (cercle élargi = 13684 thèses, dont 12035 supplémentaires non contenues dans le noyau) y fait référence dans le résumé, le titre ou les mots clefs. Une double analyse lexicale de chacun est conduite : analyse des seuls titres et mots clefs, puis analyse élargie aux contenus des résumés. Ces analyses montrent que : (1) Comme en France, il n'y a pas vraiment de marqueur commun à tous les travaux sur la FdA, il s'agit d'une juxtaposition de travaux plus ou moins thématiques. (2) Les résumés de thèses étasuniens sont beaucoup plus centrés, d'une part sur les questions de méthodologie, d'autre part sur l'individu, que leurs homologues français. (3) Chaque pays présente, d'une part des classes spécifiques, comme « didactique professionnelle » et « VAE » en France, « theology », « online learning » et « transformative learning » aux USA et d'autre part des thématiques communes : santé, langues, formation des enseignants et transformations biographiques. (4) Aux USA, on observe une forte croissance des méthodes par entretiens et une forte diminution des études quantitatives ; par ailleurs les thématiques « andragogy, instructional design » et « colleges » sont en croissance tandis que « art, social et transformative » et « health » sont en décroissance.

Mots Clefs : thèse de doctorat ; formation des adultes ; bibliométrie ; lexicométrie ; publications de recherche ; recherche aux USA

Abstract: This article explores doctoral dissertations on adult education research defended in the USA between 2000 and 2022 and accessible in the PQDT database. It completes a first study published in the previous issue of the journal *Savoirs*. It uses lexical and bibliometric analyses to propose and study two corpus of different sizes. The first one (core = 1649 theses) gathers the theses explicitly referring to the theme in their title or keywords, and the second one (enlarged circle = 13684 theses, including 12035 additional theses not contained in the core) refers to it in the abstract, the title or the keywords. A double lexical analysis of each one is conducted: analysis of titles and keywords only, then analysis extended to the contents of abstracts. These analyses show that: (1) As in France, there is not really a common marker for all the works on Adult Education, it is a juxtaposition of more or less thematic works. (2) The summaries of American theses are much more focused, on the one hand on methodological issues, and on the other hand on the individual, than their French counterparts. (3) Each country has its own specific classes, such as 'professional didactics' and 'VAE' in France, 'theology', 'online learning' and 'transformative learning' in the USA, and common themes: health, languages, teacher training and biographical transformations. (4) In the USA, there is a strong growth in interviews and a sharp decrease in quantitative studies; on the other hand, the themes of "andragogy, instructional design" and "colleges" are growing, while "art, social and transformative" and "health" are decreasing.

Keywords: research doctorate; adult education; bibliometrics; lexicometrics; research publications; research in the USA

1. Propos et contexte de cet article

Ce nouvel article de la rubrique « vie de la recherche en formation des adultes » (9eme article de la nouvelle série, repéré dans la suite comme VdR9) s'intéresse comme le précédent aux thèses de doctorat de recherche aux Etats Unis.

Dans ce dernier article (VdR8 paru dans le N° 58), nous avons effectué un premier travail de repérage et d'analyse de telles thèses soutenues aux USA pendant la dernière décennie (2010-2020) et susceptibles de traiter de questions de recherche concernant la formation des adultes. A l'occasion de ce travail, nous avons précisé le contour de la question et exploré les méthodes et données utilisables.

1.1. Rappel des apports de l'article précédent sur la complexité des doctorats aux USA

Nous avons tout d'abord constaté que les doctorats ne font pas l'objet d'un dépôt légal obligatoire aux USA et que la base de données qui permet d'en retrouver le plus grand nombre est la base commerciale ProQuest Dissertations and Thesis (PQDT, <http://proquest.com>). Celle-ci est d'ailleurs présentée par la National Library of Congress des USA (l'équivalent de la BNF en France) comme la base de référence pour le dépôt des thèses de tous les USA.

Nous avons aussi compris que -dès lors que nous voulions étudier la recherche scientifique au travers des thèses- nous devons nous limiter à ce qui aux USA est appelé « doctorats de recherche » (incluant principalement des PhD, des Ed.D. et quelques autres catégories dont les D.B.A.) en les séparant des autres types de doctorats appelés eux « doctorats de pratique professionnelle ». Nous conseillons vivement aux lecteurs intéressés par cette question de se reporter à ce précédent article.

Concrètement, nous avons étudié un premier corpus de 597 thèses (340 Ed.D., 244 Ph.D., 3 DBA et 7 autres) regroupant toutes celles incluses dans PQDT, soutenues entre 2010 et 2020 et qui contiennent au moins l'un des couples de mots (*adult ... education*) ou (*permanent ... education*) ou (*lifelong ... education*) ou (*continuing ... education*) ou (*lifelong ... learning*) connexes ou proches à deux mots près dans leurs titres ou leurs résumés. L'analyse de ce premier corpus avait montré qu'il était largement alimenté par de petites universités spécialisées et qu'il partageait des objets communs avec son équivalent français (voir articles VdR6 et VdR7, publiées respectivement dans les N°53 et 54) autour des questions de transformation identitaire, de formation des enseignants et des soignants, d'apprentissage des langues étrangères et de formation en situation de travail.

1.2. Constituer un noyau dur et un cercle élargi pour les thèses intéressant la FdA aux USA

Au cours de ce premier travail, nous avons aussi évoqué l'idée d'affiner cette première vision en reproduisant le schéma établi pour les thèses françaises, à savoir de repérer deux corpus concentriques, un noyau de thèses entièrement consacrées à des recherches sur la FdA (Formation des Adultes) et un cercle élargi, plus large, de thèses évoquant de manière substantielle des travaux de recherche sur la FdA, sans pour autant que cette thématique soit explicitement affichée dans leur titre. L'encadré 1 rappelle les raisons du choix d'un tel schéma en deux cercles concentriques ainsi que la méthode qui avait été employée dans les articles précédents sur les thèses françaises.

Encadré 1 : Observer la vie de la recherche en FdA via deux cercles concentriques de thèses

Se proposer d'analyser la vie de la recherche sur la formation des adultes en étudiant les titres et résumés des thèses revient à considérer les recherches doctorales comme un ensemble de références suffisamment balisé pour pouvoir servir d'observatoire.

Nous avons déjà signalé (article VdR6, p94) que cette méthode se heurtait à deux familles de problèmes, les premiers bibliographiques (manque d'un thésaurus consensuel sur la FdA) et les seconds plus ontologiques, prenant leurs racines « dans le fait même de prétendre catégoriser de manière manichéenne et absolue que toute thèse « serait » ou « ne serait pas » consacrée à la recherche sur la formation des adultes. Or, dès lors que l'on se met à lire les résumés, on observe qu'à contrario, nombre de thèses « intéressantes » ne sont pas à 100% consacrées aux sciences de la formation des adultes, mais comportent néanmoins une proportion significative du travail qui est éclairante sur la formation des adultes. C'est d'ailleurs sur ce fondement que, dans l'article VdR1, avaient été mises en évidence quatre familles de thèses posant problème pour la délimitation précise de la frontière du noyau dur des thèses FdA : transition vers la vie adulte ou professionnelle, développement professionnel, dispositifs multimédias et formation des enseignants.

Ce qui avait été conclu de ce constat ontologique dans VdR6 c'est qu'il était pertinent de compléter le simple noyau dur avec un « cercle élargi » de thèses (appelé à l'époque « second cercle ») beaucoup plus large que lui : il rassemblerait d'autres thèses sélectionnées non plus sur l'idée qu'elles soient à 100 % FdA mais sur le fait qu'elles approfondissent, dans certaines de leurs parties, différentes thématiques relevant de la recherche en FdA. Dans VdR6, nous avons cité l'exemple d'une thèse centrée sur les enjeux managériaux de l'informatisation de la réparation automobile qui développe une sous-investigation sur la relation entre dynamiques identitaires et rapport à la formation ; celle-ci serait repérée et positionnée au sein de ce cercle élargi et non dans le noyau.

Une méthode spécifique avait alors été développée pour constituer ce second cercle ou cercle élargi. Elle s'appuyait sur l'identification des directeurs et rapporteurs de thèses et des membres des comités scientifiques des revues françaises spécialisées dans le champ de la recherche en FdA. On en avait déduit une liste de « garants FdA » a qui servi à sélectionner une base de thèses dont les jurys étaient composés de plusieurs garants FdA qui, filtrée ensuite par proximité sémantique avec des catégories thématiques repérées dans le noyau, avait permis d'établir le corpus du cercle élargi défini plus haut.

Dans ce contexte, le propos du présent article est de chercher à constituer pour les USA un tel double corpus proposant non seulement un noyau dur de thèses explicitement focalisées sur la FdA (comme cela a déjà été fait dans VdR8) mais aussi un cercle élargi regroupant également des thèses qui approfondissent des questions de FdA sans y être pour autant à 100% consacrées. De plus, nous avons fait le choix d'étendre la période temporelle concernée pour qu'elle couvre 20 ans de thèses, dans l'idée de fournir également des éléments de références qui pourront être réutilisés prochainement dans le cadre du travail sur les 20 ans de la revue *Savoirs* (voir futur numéro 61 de cette revue).

Après cette brève introduction, cet article proposera, comme les précédents, une partie dédiée à la méthode d'identification et de récupération des thèses de doctorat susceptibles d'être concernées en tant que noyau ou cercle élargi, puis une partie consacrée à une courte présentation de leurs caractéristiques bibliographiques suivie d'une série d'analyses lexicales permettant de décrire les thématiques qu'elles traitent.

2. Définition des thèses recherchées et méthode de repérage

2.1. Méthodes de sélection : Bases et requêtes bibliographiques

A la lecture des trois derniers articles de cette rubrique, on pourrait penser qu'il suffit d'appliquer à nouveau la méthode utilisée dans VdR6 (fondée sur l'analyse de la composition de jurés, voir encadré 1) pour créer un noyau dur et un cercle élargi français à l'exploration de la partie USA de la base *Proquest Dissertations and Thesis*.

Malheureusement, force est de constater qu'une simple transposition de méthode n'est pas directement possible. En effet, la base PQDT se révèle ne pas donner suffisamment d'informations précises sur la composition des jurys pour pouvoir constituer une base fiable de « Garants FdA » aux USA. Deux problèmes se juxtaposent : la base n'est pas complètement renseignée sur la composition des jurys (près de 40% des thèses ont un champ « membre du comité » vide) et, même lorsque celle-ci est renseignée, nous ne disposons pas de liens permettant une identification univoque des jurés concernés. En effet, au vu du nombre de documents concernés (plus de 13 milliers de thèses sélectionnées dans le cercle élargi, soit environ 40 milliers de jurés) les risques de confusion d'homonymes deviennent suffisamment significatifs pour ne pas être négligés d'autant que nous ne disposons pas d'un système d'identification qui permettrait de lever les doutes (il n'y a pas de numéro IdRef ou OrCid disponible dans ProQuest PQDT).

2.2. Transformation de la méthode pour obtenir un noyau et un cercle élargi

Concrètement, il nous faut donc changer de méthode. Compte tenu des contraintes exposées plus haut, nous avons dû revenir à des requêtes plus simples se fondant sur une définition plus opératoire de la différence entre noyau et cercle élargi.

Nous avons donc adopté les deux principes suivants. (1) Si le noyau dur se caractérise par le fait que la FdA est considérée comme au cœur de la thèse par son auteur, alors la requête doit porter sur les mots du titre. (2) Complémentairement, la requête identifiant le cercle élargi doit chercher plus largement d'autres thèses dans laquelle la FdA est abordée en profondeur, mais secondairement, au service d'un autre objet de recherche (comme la professionnalisation ou la lutte contre le chômage) : pour cela la requête doit porter aussi sur le texte complet du résumé de la thèse et sur les mots clefs.

La logique peut donc être de rechercher avec la même liste d'expressions caractéristiques de la FdA, une première fois juste dans le titre des thèses pour créer le noyau dur, et une deuxième fois, cette fois dans titre, mots clefs mais aussi résumé pour créer le cercle élargi. On utilisera pour cela deux requêtes différentes dans PQDT : « TI(expression) » qui recherche les thèses contenant (expression) dans les seuls titres ou « NOFT(expression) » qui cherche en revanche dans tous les champs textuels (titre, mots, clefs, sujet et résumé) sauf le texte intégral de la thèse (NOFT = *no full text*).

2.3. Formulation des requêtes dans ProQuest Dissertation and Thesis

A l'occasion de ce changement de méthode, nous avons également décidé de compléter la grille de termes choisis pour extraire les thèses. Comme on le voit sur le tableau 1, aux termes utilisés pour VdR8 (à savoir « *adult ... education* », « *permanent... education* », « *lifelong ... education* », « *lifelong... learning* » et « *continuing... education* »), nous avons en particulier ajouté toutes les variantes des couples de termes que sont « *adult ... learner* » ou « *adult ... learning* » comme « *adults learners* » ou « *adult science learning* » qui n'étaient pas recherchées dans VdR8 ainsi que toutes les utilisations du terme « *andragogy* ».

De plus, afin de pouvoir mieux analyser d'éventuelles évolutions temporelles, nous avons décidé d'élargir la requête à toutes les thèses de doctorat postérieures à 2000, en restant néanmoins, comme déjà expliqué en VdR8 strictement limité aux thèses de doctorat de recherche comme ceux-ci sont définis dans le tableau¹ d'appartenance à la catégorie « *research doctorate degrees* » déjà utilisé dans VdR8. Le corpus est toujours limité aux thèses soutenues aux Etats Unis et écrites en langue anglaise.

¹ Ce tableau est visible à l'adresse :

<https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww2.ed.gov%2Fabout%2Foffices%2Flist%2Fous%2Finternational%2Fusnei%2Fus%2Fdoctorate.doc&wdOrigin=BROWSELINK>

Par ailleurs, il faut signaler que, depuis mai 2021, la base PQDT n'est plus que très partiellement accessible en *open access*. L'accès complet ainsi que tout téléchargement nécessitent de passer par un établissement titulaire d'un abonnement onéreux. Aussi, l'ensemble des requêtes et des téléchargements a été effectué par les auteurs au cours du mois d'août 2022 depuis une console connectée en local sur le réseau de la bibliothèque de recherche de la BNF, qui est un des rares établissements français titulaire d'un tel abonnement. De plus, les requêtes intelligentes de sortie (API ou sélection de champs de sortie) ne sont pas accessibles aux titulaires des abonnements simples à Proquest comme la BNF. Elles sont réservées à ceux qui, grâce à un investissement supplémentaire, ont également souscrit à un module ad hoc intitulé TDM (*Text and Data Mining*), malheureusement non disponible à la BNF.

Tableau 1 : Afin que l'expression de requête soit la plus efficace possible, nous avons retenu de rechercher tous les intitulés contenant **au moins l'une des expressions** présentées.

Expression recherchée	VdR8	VdR9	Requête PQDT correspondante
« Adult » suivi par « Education » à 0,1 ou 2 mots près	X	X	<i>adult PRE/2 education</i>
« Permanent » suivi par « Education » à 0 ou 1 mots près	X	X	<i>permanent PRE/1 education</i>
« Lifelong » suivi par « Education » à 0 ou 1 mots près	X	X	<i>lifelong PRE/1 education</i>
« Continuing » suivi par « Education » à 0 ou 1 mots près	X	X	<i>continuing PRE/1 education</i>
« continuing » suivi par « Learning » à 0,1 ou 2 mots près	X	X	<i>lifelong PRE/1 learning</i>
« Adult » suivi par « Train » ou « Training » ou « Trainer » à 0,1 ou 2 mots près		X	<i>adult PRE/2 train*</i>
« continuing » suivi par « Learning » à 0,1 ou 2 mots près		X	<i>adult PRE/2 learning</i>
« Adult » suivi par « Learner » à 0,1 ou 2 mots près		X	<i>adult PRE/2 learner</i>
Andragogy		X	<i>andragogy</i>

2.4. Présentation des différents corpus étudiés dans le cadre de cet article

La figure 1 présente schématiquement le positionnement des quatre classifications de lexiques qui vont être présentées dans cet article avec leurs analyses respectives : au point 3.3. les deux classifications relatives au noyau (a et b) et au point 3.5. celles relatives au cercle élargi (c et d).

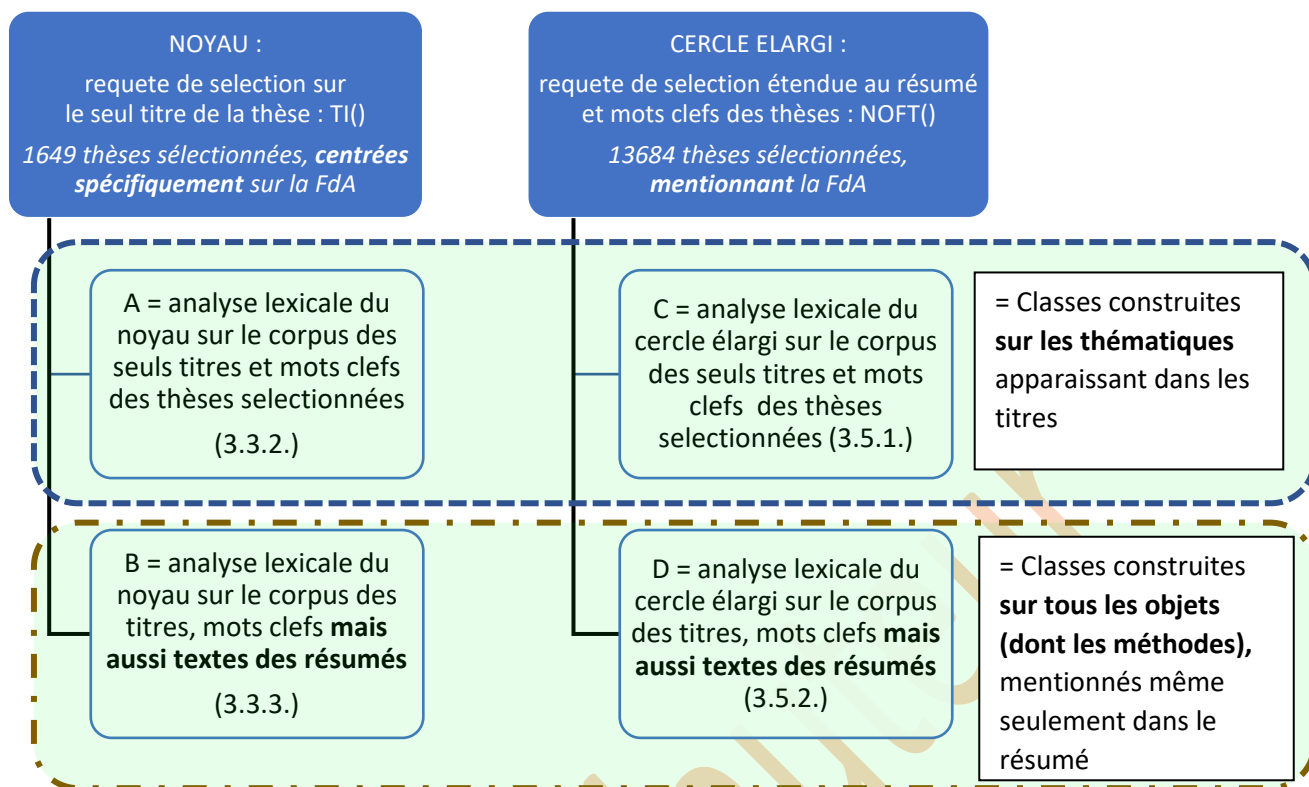


Figure 1 : complémentarités des quatre analyses lexicales présentées dans cet article.

3. Résultats des requêtes et description des thèses identifiées.

3.1. Composition du noyau et du cercle élargi

Les deux requêtes présentées ci-dessus identifient respectivement 1592 thèses dans le noyau et 13434 dans le cercle élargi (incluant par construction les 1592 du noyau). Le tableau 2 récapitule les points communs et différences entre les thèses étudiées dans VdR8 et le noyau ainsi que le cercle élargi recherché ici.

Tableau 2 : comparaison des trois sélections de thèses USA concernant la formation des adultes

	VdR8 (France)	Noyau selon VdR9	Cercle élargi selon VdR9
		Extraction VdR9 par Titre	Extraction VdR9 par NOFT
Nombre	597	1649	13684 (12035 hors noyau)
Période	2010-2020	2000-2022	2000-2022
PhD/EdD / autres	244/340/13	872/759/18	7491/6047/146
Requête posée dans la base PQDT		TI (adult PRE/2 education OR adult PRE/2 train* OR adult PRE/2 learner OR adult PRE/2 learning OR permanent PRE/1 education OR lifelong PRE/1 education OR continuing PRE/1 education OR lifelong PRE/1 learning OR andragogy) AND la.exact("English") AND SCHLOC(usa) AND pd(20000101-20221231)	NOFT (adult PRE/2 education OR adult PRE/2 train* OR adult PRE/2 learner OR adult PRE/2 learning OR permanent PRE/1 education OR lifelong PRE/1 education OR continuing PRE/1 education OR lifelong PRE/1 learning OR andragogy) AND la.exact("English") AND SCHLOC(usa) AND pd(20000101-20221231)

3.2. Comparaison du nouveau noyau avec l'échantillon étudié dans l'article VdR8

3.2.1. Nombres de thèses dans le noyau

Comme attendu, le nouveau noyau identifié par cet article regroupe plus de 2,7 fois plus de thèses que l'échantillon étudié dans l'article précédent (Vd8). Cet écart s'explique bien évidemment par le fait que la période de recueil est plus de deux fois plus longue (2000-2022 par rapport à 2010-2020), et aussi par le fait que la requête moissonne également les nouvelles terminologies, à savoir les documents mentionnant « *andragogy* » et les formes comme du type « *adult ... learning* », « *adult learner** » ou « *adult ... train** ».

3.2.2. Universités les plus productrices

Malgré ces différences, on constate des points de similitude avec l'échantillon de l'article précédent. Le premier concerne le type d'universités qui contribuent le plus à ces sélections, visible sur la figure 2. On retrouve le même établissement en tête de liste, à savoir Capella University... établissement principalement à distance et spécialisé.

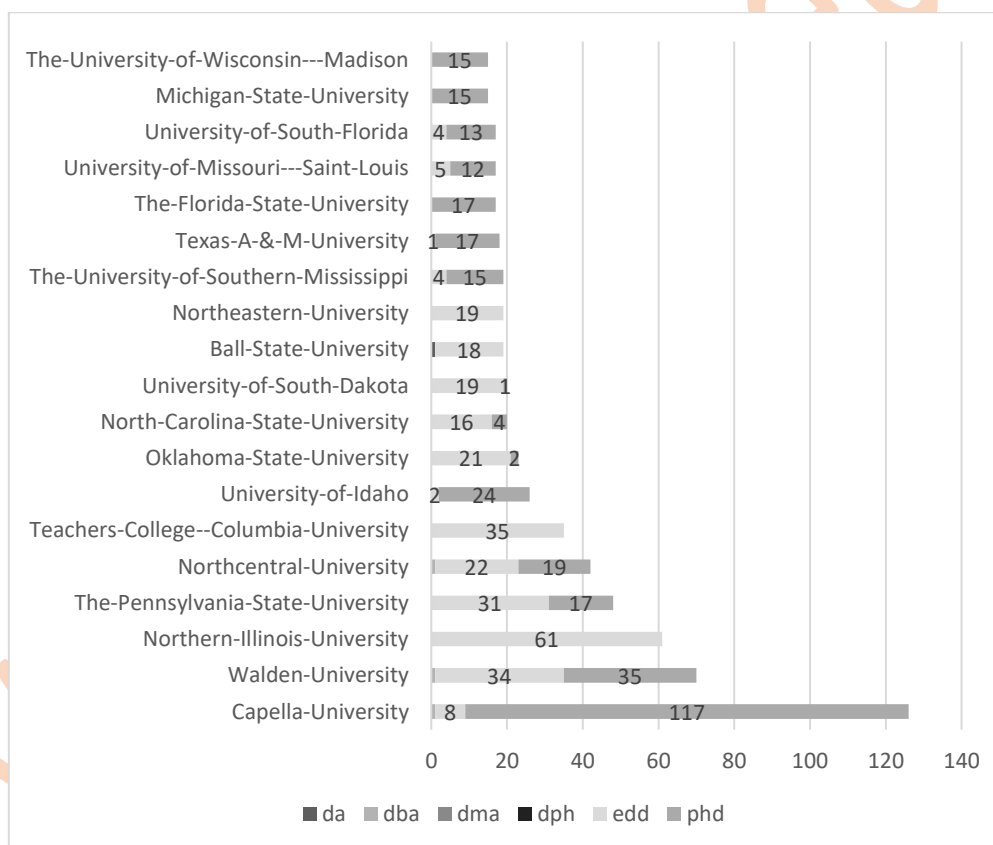


Figure 2 : Nombre de thèses du noyau par type de doctorat de recherche et par université (limité aux universités qui sont représentées dans le noyau par au moins 15 thèses)

3.2.3. Types de doctorats concernés

On voit également (figure 3) que les ratios entre les PhD, les EdD et les autres types de doctorats de recherche comme le DBA ne sont pas très différents avec une très grande majorité de PhD et EdD (seuls 1,2% n'en sont pas), même si la proportion de 55%, de EdD et 45% de PhD de VdR8 s'est inversée dans les nouvelles sélections.

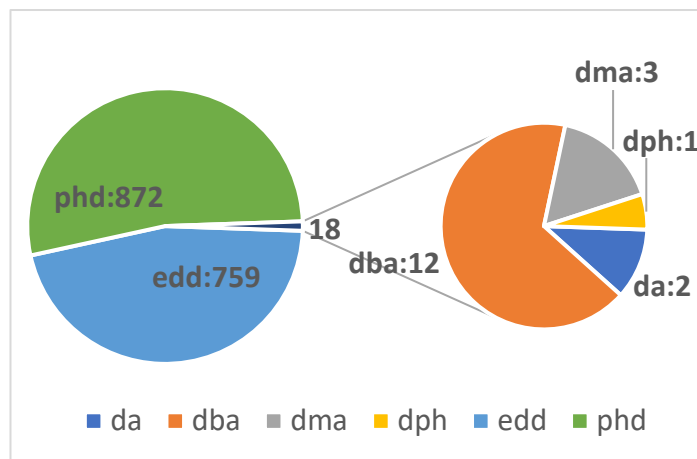


Figure 3 : répartition des thèses du noyau par type de doctorat
(da= doctorate of arts, dba = doctorate of business administration, dma= of musical arts dph= of public health, edd = education doctorate, phd =philosophical doctorate).

Sur ces deux aspects, universités concernées et type de doctorat, nous renvoyons les lecteurs aux analyses et constats déjà publiés dans l'article précédent, les résultats étant très proches malgré le doublement de la période étudiée.

3.3. Analyses lexicales des thématiques abordées dans ce nouveau noyau (1649 thèses)

3.3.1. Méthode et choix des composantes du corpus

Comme dans les articles précédents de cette rubrique vie de la recherche (VdR), le logiciel d'analyse lexicale Iramuteq a été utilisé pour cartographier les thématiques et contenus traités dans les thèses du noyau dur. Ce logiciel utilise un algorithme développé par Max Reinert (1987) qui, partant un corpus de documents, repère des groupes de mots souvent présents ensemble dans certains textes et absents des autres. Cet algorithme propose alors à partir de ces groupes de mots une double classification en quelques classes (d'une part de mots, et d'autre part des textes qui leur correspondent) permettant de repérer les principales thématiques abordées dans le corpus. Avec le logiciel Iramuteq, la classification en classes de mots est souvent présentée sous la forme d'un dendrogramme qui énumère les mots caractéristiques de chaque classe et précise le pourcentage de textes correspondant à chaque classe.

A partir du même ensemble de thèses, on peut effectuer une telle analyse lexicale selon différentes variantes. On peut par exemple prendre comme corpus de textes le corpus constitué, pour chacune des thèses, de leur résumé auquel on ajoute les titres et mots clefs (résultats d'une classification par les bibliothécaires en mots clefs). A partir du même ensemble de thèses on peut effectuer une telle analyse lexicale selon différentes variantes. On peut par exemple prendre le corpus constitué du résumé de chacune des thèses, auquel on ajoute, en tant que variante, les titres et mots clefs.

Concrètement, l'expérience montre que l'on a intérêt à effectuer ces deux variantes d'analyse et à les comparer. En effet, l'analyse menée sur les seuls titres et mots sujet crée une classification centrée sur une ou deux caractéristiques jugées comme vraiment essentielles par l'auteur et les bibliothécaires qui en ont validé les mots clefs ; de son côté, l'analyse incluant l'intégralité du résumé introduira beaucoup plus d'éléments. Les résumés des thèses soutenues aux USA semblent notamment mettre la méthodologie beaucoup plus en avant que les résumés des thèses soutenues en France. Ainsi la première analyse renseigne plutôt sur les thématiques ou objets abordés, tandis que la seconde peut la compléter en fournissant des éléments sur les méthodologies utilisées.

Dans la suite de cette présentation, nous allons présenter en détail l'analyse « titres et mots clefs » (3.3.2) et nous la compléterons par quelques commentaires issus de l'analyse incluant les résumés (3.3.3).

3.3.2. Analyse lexicale du corpus des titres et mots clefs du nouveau noyau de 1549 thèses

La figure 4 présente le dendrogramme d'une classification en 8 classes obtenue à partir du corpus composé des titres et mots clefs de ces 1649 thèses. Parmi celles-ci, le logiciel a réussi au total à classer 1464 thèses, soit 88,8% du total, dans ces 7 classes.

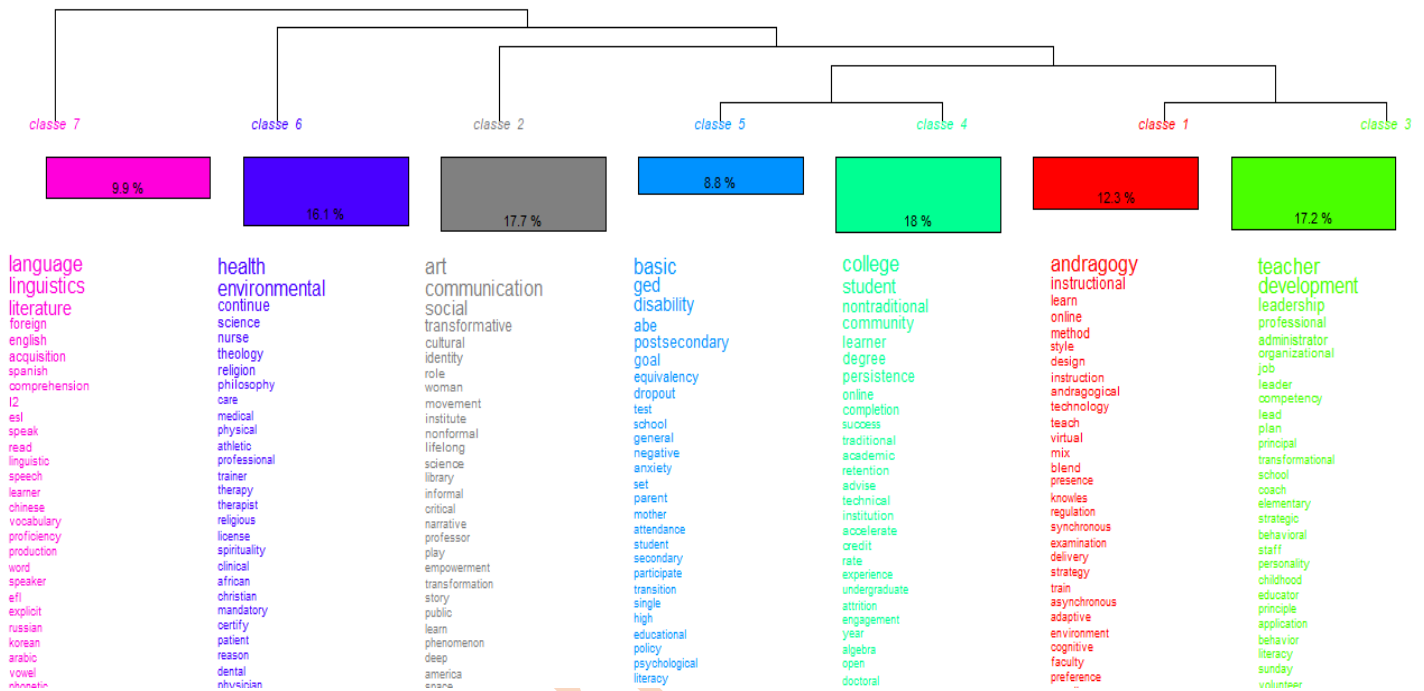


Figure 4 : Dendrogramme d'une classification en 8 classes obtenue à partir du corpus composé des titres et mots clefs des 1649 thèses du noyau.

Le logiciel Iramuteq permet également de représenter les corrélations de ces classes de lexiques avec les modalités de variables descriptives des thèses du corpus. Tous ces graphiques sont disponibles en ligne via le QR code placé en fin d'article. On y trouve donc en particulier les corrélations entre ces classes et l'année de soutenance des thèses concernées, puis les corrélations entre elles et le type de doctorat de recherche.

Dans cette analyse, les sept classes de lexiques (et de thèses) distinguées ont les caractéristiques suivantes.

- Classe a1 (180 thèses soit 13% de thèses classées) : avec comme mots les plus corrélés : « *andragogy, instructional, learn...* », cette classe regroupe les thèses traitant d'andragogie et d'*instructional design* (ingénierie pédagogique) en particulier pour les situations online. Elle se révèle en croissance sur les deux dernières décennies et regroupe une grande partie des DBA et beaucoup de PhD. La thèse la plus caractéristique de cette classe est titrée « *Andragogy and online course satisfaction : a correlation study* » (Northcentral-University, PhD, 2015).
- Classe a2 (259 thèses soit 17,7%) : avec comme mots les plus corrélés « *art, communication, social, transformative...* », cette classe regroupe les thèses s'intéressant aux champs de l'art, du

social et de la communication, en lien avec le *transformative learning*. Elle regroupe une grande partie des DA et DMA, et se trouve en décroissance sauf ces toutes dernières années). La thèse la plus caractéristique de cette classe est titrée : « *United States adult teachers and learners perspectives on representations in video games used in the classroom* » (Temple-University, PhD, 2016).

- Classe a3 (252 thèses soit 17,2%) : cette classe regroupe les thèses traitant de la formation des enseignants avec comme mots corrélés « *teacher, development, leadership, professional...* ». Il s'agit principalement d'EdD et cette classe ne montre pas une tendance d'évolution temporelle régulière sur les deux décennies. La thèse la plus caractéristique est titrée : « *leading adults through change, an action research study of the use of adult and transformational learning theory to guide professional development for teachers* » (Capella-University, PhD, 2008).
- Classe a4 (263 thèses soit 18%) : avec comme mots corrélés « *college, student, nontraditionnal...* » cette classe regroupe des thèses s'intéressant aux diplômes et étudiants des *community colleges*. Elle est en croissance régulière et regroupe beaucoup des EdD. La thèse la plus caractéristique de cette classe est titrée « *academic self efficacy and massive open online course mooc completion rates do academic self efficacy and adult learner characteristics influence mooc outcomes* » (San-Diego-State-University, EdD, 2017)
- Classe a5 (129 thèses soit 8,8%) : avec « *basic, GED, disability...* » comme mots corrélés, cette classe est centrée sur l'éducation de base des adultes et le General Education Diploma (GED). Elle regroupe une grande partie des DBA, et n'est pas marquée par une évolution temporelle régulière sur les deux dernières decennies. La thèse la plus caractéristique de cette classe est titrée : « *adults who have learning disabilities transition from GED to postsecondary activities* » (Capella-University, PhD, 2010)
- Classe a6 (236 thèses soit 16,1%) : avec « *health, environmental science, nurse...* », cette classe regroupe les thèses traitant de santé, sciences de l'environnement, infirmier ... (regroupant une grande partie des DBA, en nette décroissance). La thèse la plus caractéristique est titrée : « *factors that promote or deter trauma care nurses participation in continuing professional education* » (The-Pennsylvania-State-University, EdD, 2001)
- Classe a7 (145 thèses soit 9,9%) : « *language, linguistics, literature* ». Cette classe est corrélée aux PhD et non aux EdD et ne montre pas d'évolution temporelle régulière. La thèse la plus caractéristique est titrée : « *the effect of phonetics instruction on adult learners perception and production of l2 sounds* » (Georgetown-University, PhD, 2012)

En résumé, en termes d'évolutions sur les deux decennies, les thématiques de la classe a1 (andragogie et ingénierie pédagogique) et a4 (collèges) sont en croissance tandis que la classe a2 (art, social et transformative) et surtout la classe a6 (santé) sont en décroissance, contrairement à celle de la santé, pour ce que l'on observe en France (cf. VdR6 et 7).

3.3.3. Analyse lexicale complémentaire du nouveau noyau intégrant les résumés en plus des titres et mots clefs

Comme annoncé, nous avons également procédé à une analyse complémentaire du noyau, en incluant les textes des résumés. Ainsi, la figure 5 montre des résultats de cette nouvelle analyse du corpus

associant cette fois, pour chaque thèse, titres, mots-clés et texte des résumés ; cet ensemble de mots étant traité comme un seul texte insécable pour l'analyse lexicale. Concrètement, cela signifie que l'on a demandé à l'algorithme de re-répartir (en fonction cette fois des lexiques employés, non seulement dans les titres et mots-clés, mais aussi dans les résumés) le corpus des thèses du noyau dont 88,8% avaient été réparti lors de l'analyse précédente. Comme cela a été mentionné plus haut, cette seconde analyse -incluant les résumés- permet d'obtenir des informations plus fines, en particulier sur les méthodes mobilisées dans les thèses.

On a réglé cette fois-ci les paramètres pour obtenir une analyse en 6 classes, ce qui a permis de classer 1390 thèses soit 84.3% de l'ensemble des thèses du noyau.

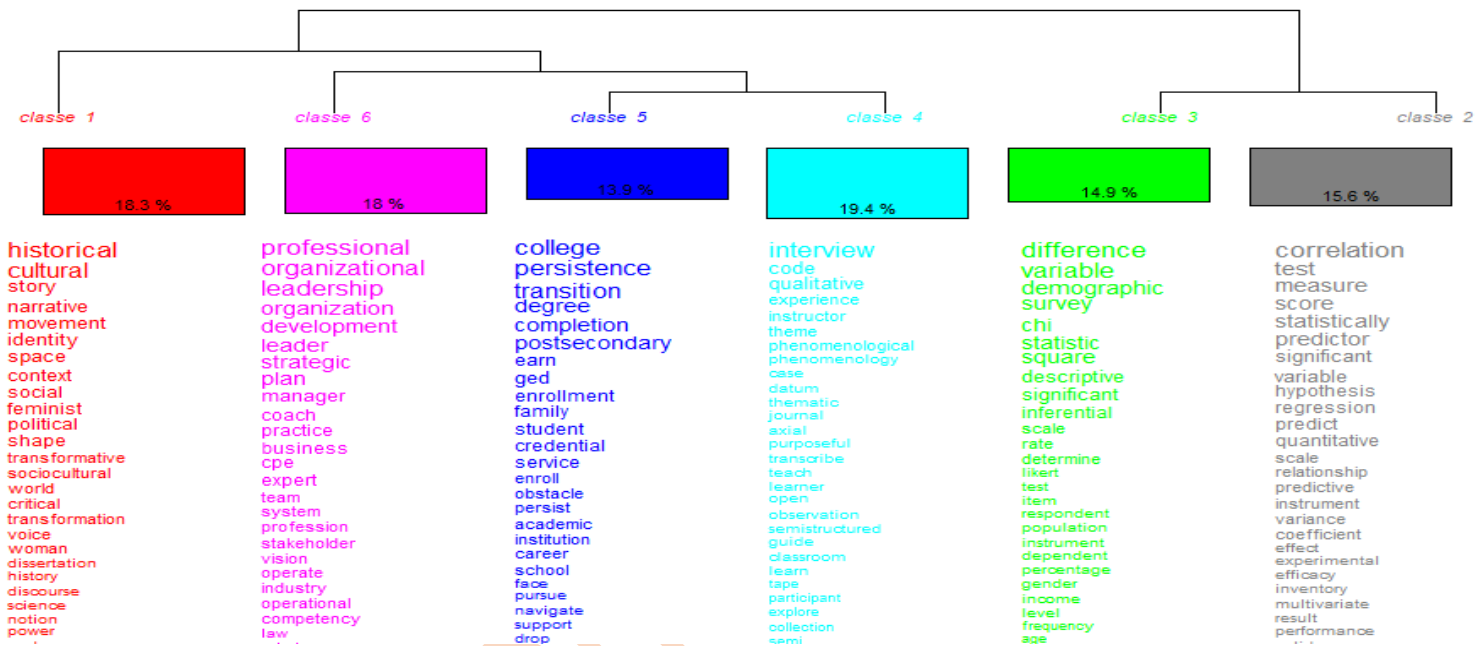


Figure 5 : Dendrogramme de la classification en 6 classes retenue à partir du corpus intégrant les résumés des 1649 thèses du noyau. Les graphiques des corrélations avec les modalités des variables sont donnés via le Q/R code en fin d'article

Comme prévu, on observe que cette analyse complémentaire met beaucoup plus au premier plan les éléments de méthode décrits dans les résumés de chacune des thèses. Ainsi, on repère maintenant certaines classes de mots (et donc de résumés+titres+mots-clés) dans lesquelles se mélangent un objet et une méthode (ou une approche) et d'autres où seules les méthodes surnagent bien au dessus des termes mentionnant divers objets, eux noyés en dessous.

De fait, trois classes correspondent à des thèses ayant en commun des méthodes.

- Classe b2 (217 thèses, soit 15,6%). Le lexique caractéristique de cette classe b2 est centré sur les études de corrélation (utilisant des méthodes de test ou d'inférence proches de la psychométrie) : « *correlation, test, mesure, score...* ». Cette classe, qui correspond à beaucoup de PhD, ne montre pas d'évolution temporelle régulière. Titre de la thèse la plus caractéristique : « *adult english learners self assessment of second language proficiency contexts and conditions* » (New-York-University, PhD, 2009).
- Classe b4 (269 thèses soit 19,4%). Le lexique de cette classe b4 est centré sur les méthodes qualitatives par interview et analyse de contenu : « *interview, code, qualitative ...* ». Cette classe b4

correspond à beaucoup des DMA et est en forte croissance temporelle sur les deux décennies. Titre de la thèse la plus caractéristique : « *understanding how adult educators apply andragogy in a correctional educational setting: a case study* » (University-of-Arkansas, EdD, 2012).

Trois classes correspondent à des associations entre des thématiques et des méthodes

- Classe b3 (207 thèses, soit 14,9%). Le lexique de cette classe b3 est centré sur des méthodes quantitatives appliquées en particulier aux variables socio-démographiques « *difference, variable, demography, survey...* ». Cette classe correspond à beaucoup de DA et DBA et est en décroissance régulière depuis 20 ans, Titre de la thèse la plus caractéristique : « *a study of the perception of students enrolled in a continuing education program at Nan Kai college Taiwan* » (University-of-Arkansas, EdD, 2004).
- Classe b1 (254 thèses, soit 18,3%) : Le lexique caractéristique de cette classe b1 associe la méthode narrative aux thématiques historico-culturelles (proches des thématiques de la précédente classe a2) : « *historical, cultural, story, narrative, social...* » ce qui indique que ces thématiques sont le plus souvent traitées par une approche narrative. Cette classe qui se retrouve dans beaucoup des DMA est en décroissance relative sauf ces toutes dernières années, comme la classe a2. Titre de la thèse la plus caractéristique : « *Beneath the surface : a narrative inquiry into educational and cultural experiences of kenyan adult women learners in USA universities* » (University-of-Wyoming, 2011, PhD).
- Classe b5 (193 thèses, soit 13,9%). Le lexique de cette classe reste thématique « *college, persistence, transition, completion* » renvoyant à la classe a4 de l'analyse précédente et qui se révèle (en retournant dans le corpus) s'appuyer le plus souvent sur des méthodes d'entretiens. Cette classe b5, tout comme son homologue a4 dans la précédente classification, correspond à beaucoup d'EdD et se révèle en forte croissance sur les deux décennies. Titre de la thèse la plus caractéristique : « *role integration of adult women online learners* » (The-Pennsylvania-State-University, EdD, 2007).
- Classe b6 (250 thèses, soit 18%). Le lexique de cette classe b6 est lui aussi thématique : « *professional, organizational, leadership* » en lien cette fois avec la classe a3, que l'on peut comprendre (en retournant également dans le corpus) comme s'appuyant sur de l'analyse de pratiques ou des organisations. Cette classe b6, tout comme son homologue a3, correspond à beaucoup d'EdD et ne montre pas d'évolution temporelle régulière sur les 2 décennies. Titre de la thèse la plus caractéristique : « *exploring programmatic issues which affect continuing legal education practice in Kansas* » (Kansas-State-University, EdD, 2017).

Conclusion sur les évolutions des méthodes dans les thèses du noyau

En synthèse, parmi les résultats marquants en termes d'évolution des méthodes mobilisées, on voit se dégager, pour ce corpus du noyau (voir tableau 3), des évolutions de méthodes sur ces derniers 20 années. Une montée en puissance des classes b4 et b5 (études « phénoménologiques » par entretiens) et une forte décroissance de la classe b3 des études quantitatives d'inspiration sociodémographique.

Même si cette évolution est très frappante, on ne peut néanmoins pas conclure à une régression totale du quantitatif au profit du qualitatif car on a assisté dans le même temps à une décroissance de la classe b1 sauf ces dernières années (qualitative : narrativité, historico-culturelle). De plus, on observe

l'existence en parallèle avec la classe b3 d'une autre classe « quantitative » b2 (qui elle n'est ni en croissance ni en décroissance) et qui correspond à des études de corrélation de type psychométrique.

Tableau 3 : Evolution temporelle des méthodes développées dans les résumés des thèses du noyau

Méthodes	Etudes narratives	Etudes quantitatives par questionnaires ou échelles		Etudes qualitatives par entretiens		Analyse de pratiques ou des organisations
Classe	b1 (a2)	b2	b3	b4	b5 (a4)	b6
Description de la classe	Etudes narratives historiques, culturelles ou sociales	Etudes de corrélation (test ou d'inférence proches de la psychométrie)	Etudes quantifiées appliquées aux variables socio-démographiques	Etudes phénoménologiques par entretiens	Etudes de persistance et de complétion dans les « colleges »	Etudes sur le leadership (organisationnel ou professionnel)
Thématique dominante le cas échéant	<i>Historical, Cultural</i>	Pas de thématique dominante repérée	Pas de thématique dominante repérée	Pas de thématique dominante repérée	<i>Colleges persistence and completion</i>	<i>Leadership</i>
Evolution temporelle	Décroissance sauf ces dernières années	Pas d'évolution régulière	Forte décroissance	Forte croissance	Forte croissance	Pas d'évolution régulière
Doctorat le plus corrélé	DMA	PhD	DA et DBA	DMA	EdD	EdD

De plus, pour ce qui est des évolutions temporelles des thématiques, le rapprochement de l'analyse par les résumés (classes « b ») avec celle par titres (classes « a ») confirme deux évolutions importantes : celles des classes « cultural » (repérée en décroissance à la fois comme a2 et b1) et des classes « college » (repérée en forte croissance à la fois comme a4 et b5).

3.3.4. Relation avec l'analyse de l'article précédent (VdR8)

Même si le corpus est bien plus large (période et requête), il est intéressant de comparer ces nouvelles classifications avec celle publiée dans VdR8 portant sur le noyau alors identifié (597 thèses). Le lecteur assidu pourra vérifier que les classes trouvées dans cet ancien noyau peuvent toutes être liées aux classes du nouveau², certaines à celles de la nouvelle analyse des titres et mots-clefs (classes « a ») et d'autres à celle de l'analyse incorporant les textes des résumés (classes « b »).

3.4. Description du cercle élargi (au total en incluant le noyau 13684 thèses)

Pour éviter toute ambiguïté, nous rappelons que ce « cercle élargi » contient les 13684 thèses obtenues par la requête NOFT(...) du tableau 2. Ainsi, par construction il contient en particulier l'ensemble des 1649 thèses du noyau décrites précédemment. Dans la suite de l'article, ces 1649 thèses n'appartenant qu'au noyau seront notées « TIT » (puisque la requête qui les sélectionne recherche des termes dans les titres seuls) et les 13684-1649 n'appartenant pas au noyau mais appartenant au cercle élargi seront notées « NOTIT ».

Les figures 6 et 7 présentent respectivement la composition de ce cercle élargi (et en particulier du corpus de thèses NOTIT) en termes d'universités de soutenance et de type de doctorats.

² On retrouve ainsi les relations suivantes entre les six classes mises en évidence dans VdR8 : *teacher, interview, district* = A3 (en partie B4) ; *College, Regression* = A4, B5 (et en partie B3) ; *Language* = A7 (noyé dans les B dont le B4 =interview, theme) ; *Employee, Organization* = B6 ; *Transformative, Story* = A2 (et en partie B1) ; *Score, Difference, Significant* = B2

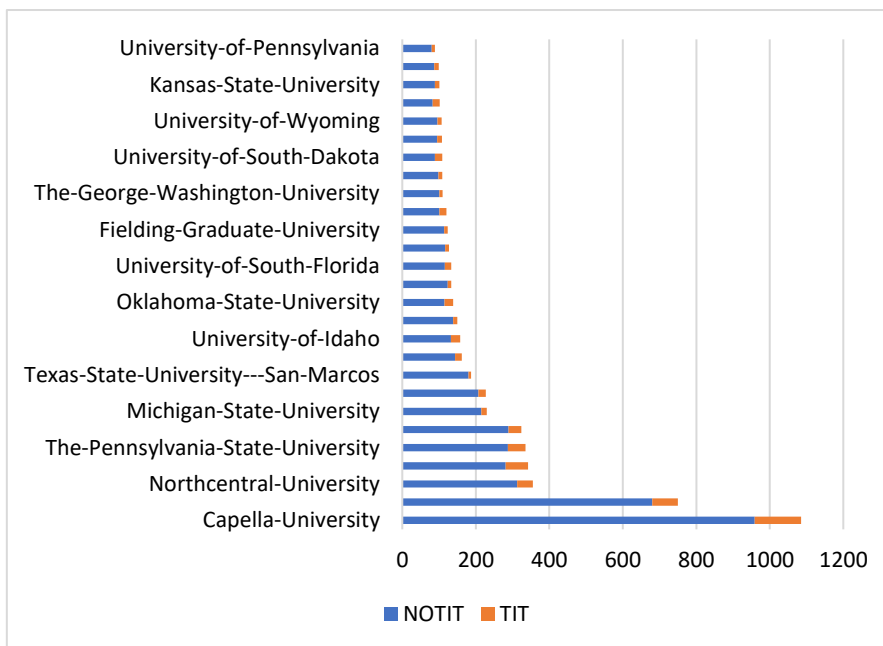


Figure 6: Universités les plus contributives en nombre de thèses trouvées dans le cercle élargi, soit dans le noyau (TIT) soit à l'extérieur du noyau (NOTIT). Cette figure est limitée aux universités contribuant avec plus de 88 thèses sur la période.

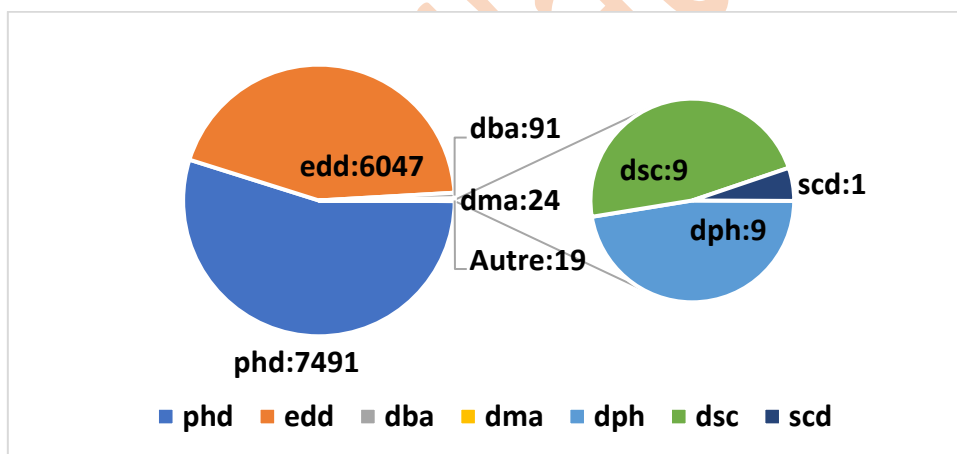


Figure 7 : Répartition des thèses NOTIT (cercle élargi mais non noyau) par type de doctorat

3.5. Analyses lexicales des thèses du cercle élargi

Comme pour l'analyse du noyau, deux variantes de l'analyse du cercle élargi sont proposées. La première étudie le corpus constitué par les seuls titres et mots clefs dans le but de décrire les grandes thématiques traitées ; la seconde étudie le corpus incluant aussi pour chaque thèse son résumé de façon à pouvoir aussi décrire, le cas échéant, les méthodologies en jeu. Ces analyses seront toujours effectuées en utilisant l'algorithme Reinert implanté dans le logiciel Iramuteq.

3.5.1. Analyse lexicale des seuls titres et mots clefs du cercle élargi

Pour cette première variante, la figure 8 présente le dendrogramme d'une analyse en 8 classes qui a permis de classer 86% des thèses du corpus.

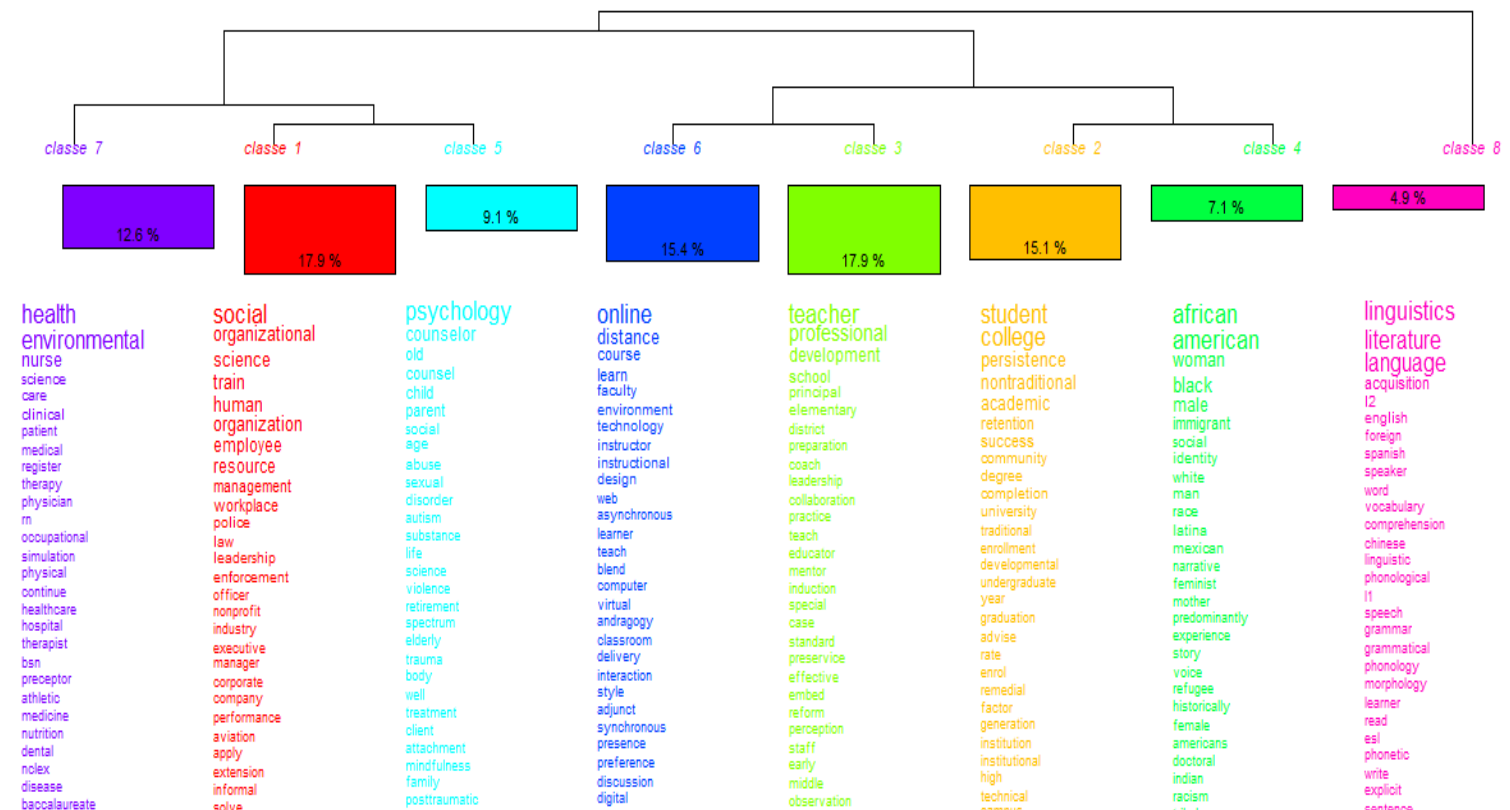


Figure 8 : Dendrogramme d'une classification en 8 classes, obtenue à partir du corpus composé de titres et mots clefs des 13684 thèses du cercle élargi. Les graphiques des corrélations avec les modalités des variables sont donnés via le QR code en fin d'article

Quatre d'entre elles se révèlent être de simples prolongements de classes déjà repérées dans le noyau :

- Classe c2 (1783 thèses soit 15,1%) : Cette classe centrée sur « *student, college, persistence...* » est un prolongement des classes a4 et b5 du noyau. Comme dans le noyau, on observe une corrélation avec le type EdD ; en revanche elle n'est pas ici en croissance, sans doute par ce qu'elle mélange aux études qualitatives (comme celles qui composaient les classes a4 et b5) d'autres études cette fois quantitatives. Titre de la thèse la plus caractéristique : « *Self-directed learning characteristics of first generation first year college students participating in a summer bridge program* » (2011, EdD, University-of-South-Florida)
- Classe c3 (2106 thèses soit 17,9%) : Cette classe « *teacher, professional, développement...* » est un prolongement de la classe a3. Elle est ici toujours corrélée avec la catégorie EdD, en revanche ici en forte croissance. Titre de la thèse la plus caractéristique : « *Effects of LTTA technology training on the perceptions of Nebraska school principals about their ability to satisfy technology skills and national teacher and administrator standards* » (2003, EdD, The-University-of-Nebraska---Lincoln)
- Classe c7 (1483 thèses soit 12,6%) : Cette classe, centrée sur « *health, environmental, nurse...* », est très proche de la classe a6, mais cette fois ci, il n'y a pas d'évolution temporelle régulière. Titre de la thèse la plus caractéristique : « *Content analysis of applied learning from high fidelity patient simulation orientation to critical care* » (2015, PhD, City-University-of-New-York)

- Classe c8 (578 thèses soit 4,9%) : Cette classe, centrée sur « *linguistics, literature, language...* », est très proche de la classe a7 (en décroissance, correspondant plus à des PhD). Titre de la thèse la plus caractéristique : « *L2 and L3 acquisition of the portuguese stressed vowel inventory by native speakers of english* » (2011, PhD, The-University-of-Arizona)

Les quatre autres classes correspondent à des thématiques non relevées dans le noyau :

- Classe c1 (2116 thèses soit 17,9%) : « *social sciences, organizational, train ...* » ; cette classe, regroupant des DBA et des PhD, est en forte décroissance temporelle. Titre de la thèse la plus caractéristique : « *Factors influencing the interagency collaborative behavior of North Carolina field employees in public sector non regulatory natural resources management* » (2017, EdD, North-Carolina-State-University)
- Classe c4 (839 thèses soit 7,1%) : « *african, american, woman, black...* ». Cette classe, en partie en lien avec a2 (“narrative”) et peut-être a5, est en croissance ces dernières années. Titre de la thèse la plus caractéristique : « *Navigating the path toward graduation : a qualitative case study of african american male persistence at a predominantly white institution* » (2013, EdD, The-University-of-Memphis).
- Classe c5 (1069 thèses soit 9,1%) : « *psychology, counsellor, old, counsel...* » ; elle regroupe surtout des PhD. Titre de la thèse la plus caractéristique : « *The voices of female mental health professionals working with american indians with co occurring disorders: a constructivist grounded theory* » (2013, PhD, The-University-of-Texas-at-San-Antonio)
- Classe c6 (1816 thèses soit 15,4%) : « *online, distance, course, learn...* » ; pas d'évolution temporelle régulière. Titre de la thèse la plus caractéristique : « *Building learning communities facilitating interaction in computer mediated online courses* » (2008, PhD, University-of-Wyoming)

Rosace de représentation de la structure des classes du cercle élargi

Concrètement, ouvrir la recherche des thèses au-delà du noyau s'est donc traduit par trois phénomènes distincts : (1) le nombre de thèses faisant partie des quatre classes déjà repérées dans le noyau est quasiment décuplé (a4/c2 = student college, a3/c3 = teacher development, a7/c8 = language, a6/c7 = health). Elle passent d'environ deux centaines à deux milliers dans chacune des classes ; (2) quatre nouvelles classes thématiques sont apparues (c1, c4, c5 et c6) ; (3) trois classes du noyau se sont diluées dans son élargissement, à savoir a1 = *andragogy*, a2 = *art social*, et a5 = basic, GED, dropout.

Pour mieux visualiser ces trois phénomènes, la figure 9 propose une représentation symbolique des relations entre les deux classifications portant sur les titres et mots clefs (« a » pour celle des thèses du noyau et « c » pour celle du cercle élargi). Sur cette figure 9, les quatre classes « c » qui sont de simples extensions des classes « a » du noyau sont figurées pour simplifier sous la forme d'une inclusion totale de leurs équivalents dans le noyau.

En revanche, lorsqu'une classe « a » n'est pas entièrement incluse dans une classe « c », les délimitations des intersections entre ces classes « a » et « c » n'ont été tracées que de manière simplement indicative : nous avons représenté par exemple le fait que a1 avait en commun le terme « online » avec c6, que a2 avec c1 avaient une recouvrement grâce au terme « social », que c1 n'avait sans doute pas une intersection vide avec a1 et enfin que a5 pouvait avoir un lien non vide avec c4 et c5 (publics spécifiques et conseil en orientation). Néanmoins, bien d'autres chevauchements sont possibles

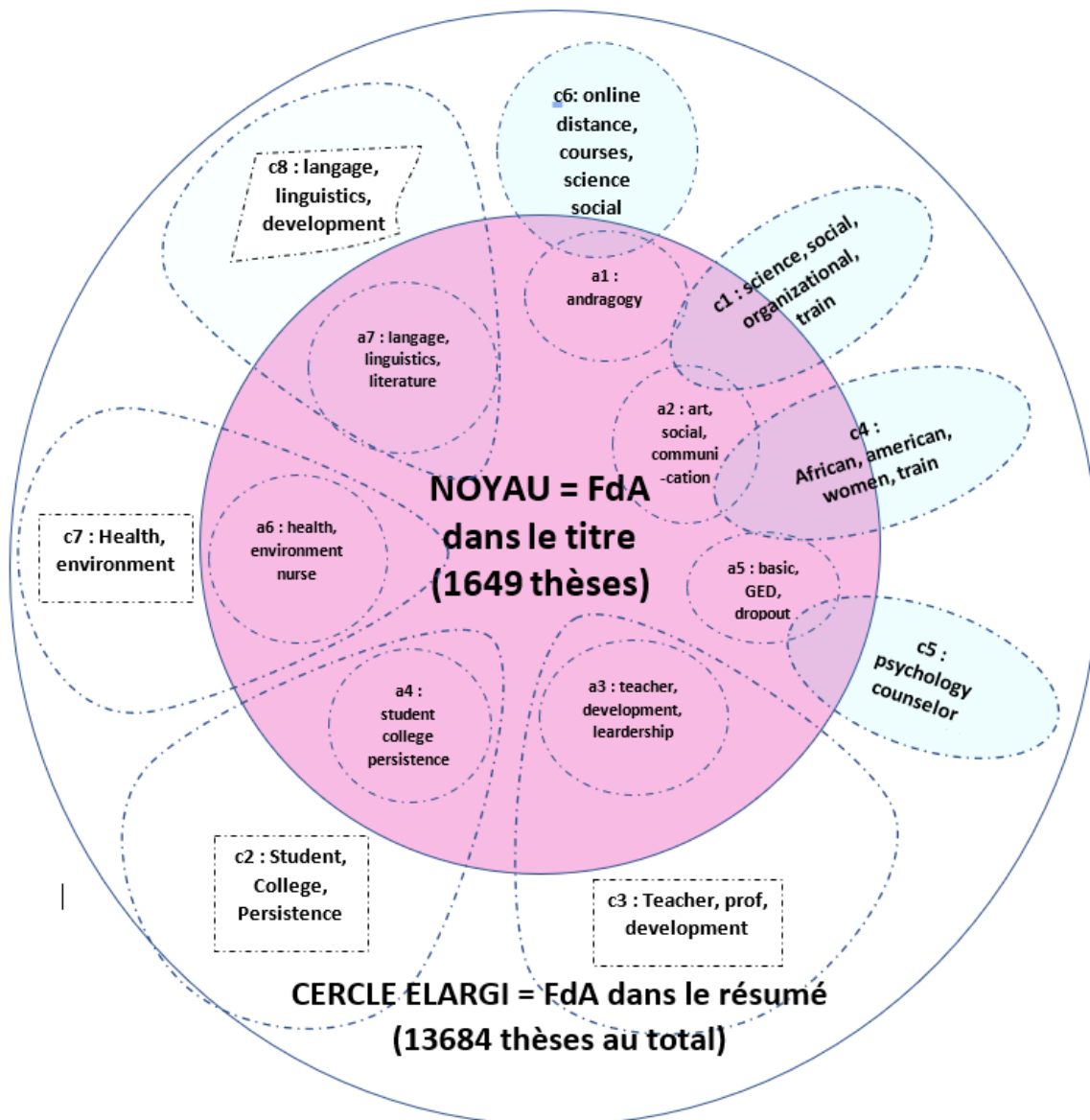


Figure 9 : décomposition symbolique des classes de titres et mots clés du noyau « a » et du cercle élargi « c » composant le cercle élargi. Lorsqu'une classe « a » n'est pas entièrement incluse dans une classe « c », leurs intersections n'ont été tracées que de manière indicative (voir texte).

3.5.2. Analyse secondaire du cercle élargi en prenant en compte les résumés.

Comme pour l'analyse du noyau, nous allons regarder ce que donne une analyse secondaire du corpus du cercle élargi, intégrant cette fois pour chaque thèse, en plus des titres et mots-clés, les résumés.

Avec cette analyse complémentaire (figure 10), on retrouve, comme dans le cas de l'analyse secondaire du noyau, des classes centrées sur des méthodes et d'autres correspondant aux quelques familles thématiques surnageant des résumés.

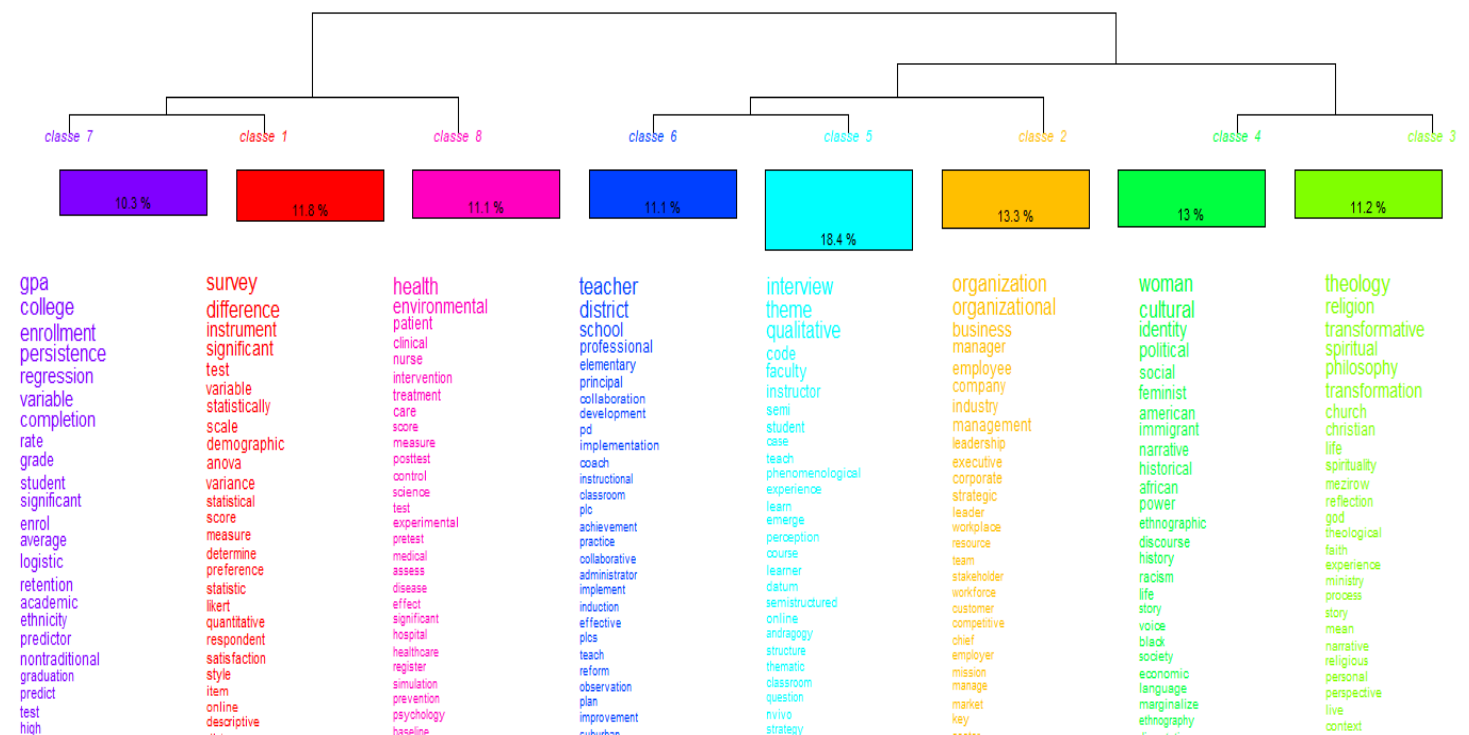


Figure 10 : Dendrogramme de la classification en 8 classes, retenue à partir du corpus intégrant les résumés des 13684 thèses du cercle élargi. Les graphiques des corrélations avec les modalités des variables sont donnés via le QR code en fin d'article

Deux classes liées exclusivement à des méthodes, l'une quantitative (en forte décroissance) et l'autre qualitative, (en forte croissance) apparaissent :

- Classe d1 (1343 thèses soit 11,8%). Avec « *survey, difference, instrument ...* », il s'agit de la prolongation de la classe de méthode quantitative b3 repérée dans le noyau. Elle est toujours en forte décroissance temporelle. Titre de la thèse la plus caractéristique : « *A study of the impact of instructor behaviors and teaching methodologies on student success in internet based courses at a South Texas college* » (2006, EdD, Texas-A-&-M-University-Kingsville)
- Classe d5 (2102 thèses soit 18,4%). Avec « *Interview, theme, qualitative...* », il s'agit symétriquement de la prolongation de la classe de méthode qualitative b4 repérée dans le noyau. Cette classe est dans les deux périmètres en forte croissance. Titre de la thèse la plus caractéristique : « *Faculty preparedness for transition to teaching online courses in the Iowa community college online consortium* » (2015, phd, Iowa-State-University).

Les autres six classes se révèlent elles encore liées à des thématiques de contenu :

- Classe d2 (1514 thèses soit 13,3%) : *organization, organizational, business ...* Cette classe est en relation avec la classe de thématique c1 du cercle élargi et b1 des résumés du noyau. Titre de la thèse la plus caractéristique : « *Health care reform leading successfully through reform rapids* » (2018, phd, University-of-St-Thomas—Minnesota).
- Classe d3 (1274 thèses soit 11,2%) : « *theology, religion, transformative ...* ». Cette classe peut être mise en relation avec la classe a2 qui fait aussi référence au « *transformative learning* » mais elle la regroupe avec le champ de la théologie. Titre de la thèse la plus caractéristique : « *Closing the sunday monday gap : vocational transformation narratives of christian young adults* » (2019, phd, U_Michigan-State-University).

- Classe d4 (1480 thèses soit 13%) : « *woman, cultural, identity ...* ». Cette classe qui englobe c4 semble centrée sur les inégalités avec une logique sous-jacente d'intersectionnalité. Titre de la thèse la plus caractéristique : « *Unraveling a lifetime of racism and sexism: an autoethnography of the educational journey from kindergarten to doctoral education for an african american woman* » (2020, edd, Northern-Illinois-University).
- Classe d6 (1261 thèses soit 11,1%) : « *teacher, district, school ...* ». Cette classe concerne le développement professionnel des enseignants : elle est très proche des classes a3 et c3 mais en regardant plus précisément le corpus on retrouve de nombreuses mentions de « communautés professionnelles d'apprentissages » (en américain : « *plc* » et « *plcs* »). Elle est en croissance. Titre de la thèse la plus caractéristique : « *collaborative leadership practices in middle schools to improve student learning in english language, arts and math* » (2017, edd, Sage-Graduate-School)
- Classe d7 (1171 thèses soit 10,3%) : « *GPA³, college, enrollment, persistence...* ». Il s'agit d'une classe qui semble très proche de c2, et on peut confirmer qu'elle mélange aux études qualitatives d'autres études cette fois quantitatives avec des mots comme « *regression, variable, significant* ». Comme c2, elle n'est pas ici en croissance. Titre de la thèse la plus caractéristique : « *understanding student success and institutional outcomes in service learning coursework at a North Carolina community college : a propensity score study* » (2016, phd, North-Carolina-State-University).
- Classe d8 (1268 thèses soit 11,1%). Avec « *health, environment, patient...* », il s'agit d'une classe liée à la thématique de la santé correspondant à c7 (et donc aussi a6). Titre de la thèse la plus caractéristique : « *A preliminary efficacy and feasibility of an obstructive sleep apnea educational intervention in Oman* » (2017, phd, George-Mason-University).

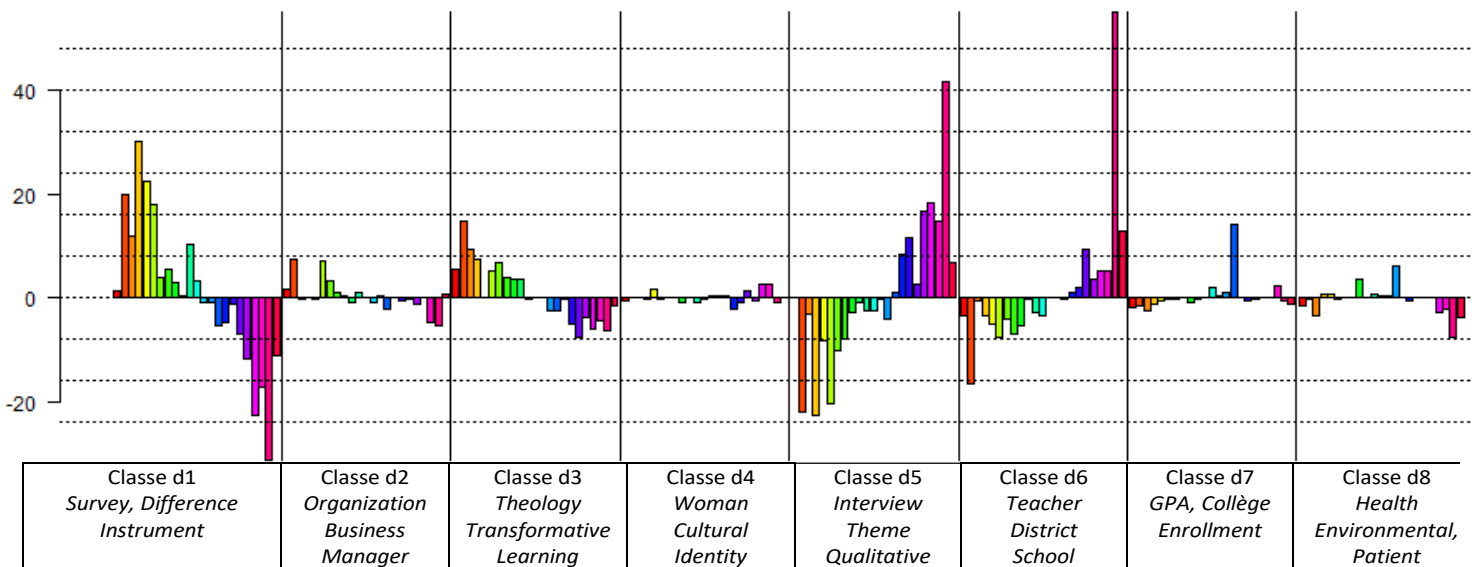


Figure 11 : correlations (ou anti correlations) entre les classes « d » et les années de soutenance. Chacun des 8 histogrammes va de 2000 à 2021 (année par année)

³ Utilisé par les *high schools, colleges, and graduate schools*, le GPA (*Grade Point Average*) est un chiffre unique et cumulatif qui représente l'ensemble des résultats scolaires.

3.5.3. Synthèse des analyses du noyau et du cercle élargi

Constats sur les thématiques :

Comme on a pu le figurer sur la rosace de la figure 9, l'analyse des titres seuls du cercle élargi confirme les quatre grandes thématiques principales déjà visibles dans le noyau que sont :

- *Teacher, professional, développement* (252 dans le noyau par comparaison avec 2106 dans le cercle élargi)
- *Student, college* (263 dans le noyau par comparaison avec 1783 dans le cercle élargi)
- *Health, environmental, nurse* (236 dans le noyau, 1483 dans le cercle élargi)
- *Linguistics, language* (145 dans le noyau, 578 dans le cercle élargi)

En revanche les trois autres classes de titres du noyau (*andragogy, art* et *social* ainsi que *basic, GED*) se sont diluées parmi les autres thèses du cercle élargi et leur groupe a disparu dans l'analyse, mais quatre autres classes ont vu le jour :

- *online learning* (pas d'évolution temporelle notable, 1816 thèses dans le cercle élargi)
- *afro american woman learners* (en croissance temporelle, 839 thèses dans le cercle élargi)
- *psychology counselling* (pas d'évolution temporelle notable, 1069 thèses dans le cercle élargi)
- *social, science, organization* (en forte décroissance temporelle, 2116 thèses dans le cercle élargi)

Et l'analyse des résumés + titres du même cercle élargi (d) a ajouté une cinquième thématique supplémentaire :

- *theology, religion, transformative learning* (en décroissance temporelle)

Le fait que de nouvelles classes apparaissent n'a rien de surprenant. En effet, le cercle élargi contient par définition des thèses dont l'objet principal n'est pas la FdA mais qui en parle à l'occasion d'un travail de recherche sur autre chose que la FdA elle-même. En revanche, ce qui peut retenir l'attention ce sont les thématiques.

(1) Il est étonnant que les classes « *on line learning* » et « *social, science, organization* » ne soient pas déjà des classes thématiques visibles dans le noyau car on pourrait penser qu'il s'agit de champs fondamentaux pour la formation des adultes dans tous les pays.

(2) Symétriquement, il est aussi inattendu que n'apparaissent pas des thématiques « secteur » ou « métiers » ou encore « contenus » comme « ingénieur » ou « computer science » par exemple. Cela signifie -en creux- qu'au-delà des métiers de l'enseignement ou de la santé et des contenus linguistiques, il n'existe pas aux USA de grand domaine dans lequel la recherche en FdA aurait été développée dans les dernières décennies. Enfin, au même titre que la classe intersectionnelle « *afro, american, woman* », on aurait pu aussi imaginer une ou plusieurs autres classes de bénéficiaires spécifiques comme par exemple « *veterans* ».

Constats concernant les méthodes

C'est sur le plan des méthodes que les résultats de ces analyses sont les plus frappants. On observe plusieurs phénomènes qui prolongent et amplifient ceux déjà signalés sur le noyau :

(1) Les descriptions des méthodes employées sont tellement détaillées dans les résumés qu'elles donnent naissance à des classes de thèses spécifiques.

(2) Toutes les analyses montrent une forte décroissance des études quantitatives (voir d1 par exemple) ; en revanche la croissance des études par entretiens est très forte, aussi bien en général (d5,

b4) que pour les études de persistance dans les collèges (a4/b5). Ce phénomène est particulièrement bien visible sur le tableau 3.

(3) Même si les deux décennies ont vu la montée en puissance des études qualitatives, la situation pour ce qui est des études narratives (histoire de vie, biographisation) est différente : leur nombre a décliné pendant les 15 premières années de la période, mais elles se redressent depuis.

4. Comparaison avec les analyses des thèses soutenues et déposées en France

4.1 Effectifs des corpus, ratios de dépôts de thèses et longueur des résumés des thèses

Une première comparaison peut être faite sur le volume de thèses de doctorat de recherche intéressant la formation des adultes et déposées dans les bases de référence. Le tableau 4 montre que ce ratio de thèses par habitant est à du même ordre de grandeur aux USA qu'en France, soit environ une par million d'habitants tous les quatre ans pour ce qui concerne le noyau et presque une par an et par millions d'habitants pour le cercle élargi. Attention néanmoins car nous ne savons pas (voir VdR8) quel pourcentage des thèses soutenues aux USA est vraiment déposé dans PQDT.

Tableau 4 : Ratios de thèses de doctorat de recherche intéressant la formation des adultes en France et aux USA déposés dans les bases ad hoc (Sudoc en France et PQDT aux USA).

	France sur 10 ans	Extrapolation sur 21 ans	France par an	USA sur 21 ans	USA par an	France par Million d'habitants et par an (67 millions)	USA par Million d'habitants et par an (331 millions)
Noyau	160	336	16	1649	78	0.23	0.24
Cercle élargi	619	1300	61	13684	224	0.91	0.67
Ratio CE/N	3.81		2.87		3.81		2.87

Toujours d'un simple point de vue de comptage, on peut aussi remarquer que les résumés des thèses aux USA sont un peu plus longs qu'en France (1,1 fois en nombre médian de caractères⁴ alors que de manière générale la langue anglaise est considérée comme plus compacte d'un facteur 0,75).

4.2 Comparaison du noyau USA avec le noyau des thèses soutenues en France

Pour disposer d'une première vision d'ensemble des différences et similitudes France/USA, on peut s'appuyer sur la visualisation par « arbre maximum des similitudes » qui est développée dans le logiciel Iramuteq.

Cette visualisation fait apparaître sous forme d'un arbre les co-occurrences deux à deux qui existent entre les mots dans chaque texte composant un corpus. Les mots sont d'autant plus gros qu'ils sont fréquents. L'épaisseur du trait liant deux mots augmente avec le nombre de textes où ces deux mots sont en co-occurrence. Ne sont tracés que les traits liant deux mots sans créer de cycle fermé (Verges et Bourriche, 2001 ; Ratinaud et Marchand, 2012). Les figures 13 et 14 présentent les résultats de ces visualisations pour les corpus des thèses des noyaux avec résumé (USA et France).

⁴ Longueur médiane des résumés en caractères espaces compris : 1765 pour la France et 1946 pour les USA.



Figure 12 : arbre maximum des similitudes du noyau USA avec résumé, limité aux mots apparaissant plus de 300 fois dans le corpus et aux co-occurrences supérieures à 150 fois

La comparaison des deux figures permet d'observer des spécificités entre les deux corpus⁵ :

En ce qui concerne le corpus USA, on voit en particulier : (1) une centration sur l'individu avec le rôle central du mot « *adult* » (2) le poids des méthodes, marqué par l'existence d'une branche très feuillue « *study* » puis « *datum* », (3) la présence importante d'une sémantique « *education* ».

En ce qui concerne le corpus France, en plus d'une centration clairement positionnée au niveau meso avec le rôle central du mot « *formation* », on retrouve l'existence d'une branche enchaînant « *professionnel* » et « *activité* » (correspondant entre autres à la spécificité française de la didactique professionnelle et des analyses de l'activité.

4.3. Comparaison avec cercle élargi en France

Pour analyser maintenant plus finement les différences et les similitudes entre les classes construites ici aux USA avec celles construites antérieurement (VdR6 et VdR7) pour la France, on peut par exemple faire la comparaison France/USA entre des classifications issues des cercles élargis avec résumé. Cela revient à comparer la classification USA de la figure 10 avec son équivalent sur les thèses françaises (publié dans VdR7) telle qu'il est reproduit figure 14.

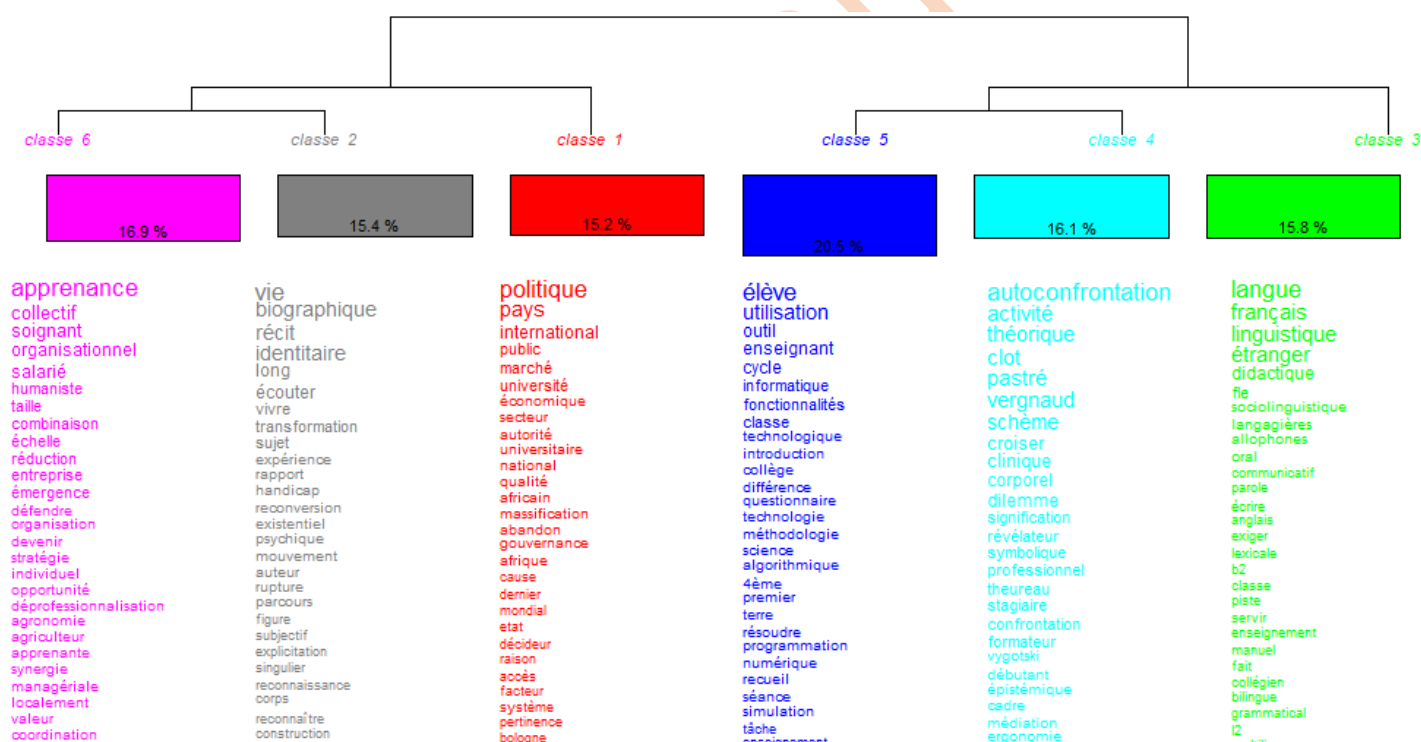


Figure 14 : dendrogramme de l'analyse avec les résumés du cercle élargi en France (source : VdR7).

Tout d'abord, on observe bien que contrairement aux USA, nous avons à faire ici à des classes thématiques et non méthodologiques et qu'il y a juste une classe intermédiaire : « analyse de l'activité par autoconfrontation ». Cela confirme que les résumés français sont beaucoup moins orientés méthode qu'aux USA, sans pour autant présager de cette tendance dans les thèses elles-mêmes.

⁵ Une analyse comparative similaire faite avec les deux corpus des cercles élargis montre que certaines de ces tendances sont toujours visibles (spécificités françaises de la VAE et de la didactique pro), mais que le centrage USA sur les « *adults* » disparaît, noyé dans l'ensemble des thèses. L'importance consacrée aux méthodes dans les résumés reste aussi parfaitement visible dans les analyses du cercle élargi des USA par comparaison à la France.

Une fois ce constat fait, on peut détailler les correspondances entre les deux classifications comme cela est fait dans le tableau 5.

Tableau 5 : comparaison des classifications des deux cercles élargis (France et USA)

N° de classe en France	Classe Fd1 « politique, pays international »	Classe Fd2 (15 %) : « vie, biographie, récit »	Classe Fd3 (16 %) : « langue, français, linguistique »	Classe Fd4 (16 %) : « auto-confrontation, activité, théorique »	Classe Fd5 (21 %) : « élève, utilisation, outil, enseignants »	Classe Fd6 (17 %) : « apprenance, collectif, soignant »
Caractéristiques des thèses concernées en France	Thèses centrées sur les politiques de formation en particulier économiques	Thèses centrées sur les histoires de vie et la biographisation	Thèses centrées sur le FLE et les langues en général	Thèses centrées sur l'analyse de l'activité et la didactique professionnelle	Thèses centrées sur le système scolaire et la formation des maîtres	Thèses concernant la motivation ainsi que les travaux liés aux professionnels de santé
N° de Classe aux USA	PAS D'EQUIVALENCE	D3 et D4	C8	PAS D'EQUIVALENCE	D6 et D7	En partie D8
Correspondances dans les thèses aux USA	Ce thème ne constitue pas une classe aux USA : pas de groupe de thèses ayant cette spécificité : peut-être à mettre en relation avec l'absence de politique fédérale de FdA, contrairement à la France	Cette thématique est mobilisée aux USA dans les deux classes d3 et d4 (Transformative et Woman)	Cette thématique est très visible dans la classe c8 mais est ensuite diluée par l'ajout des textes des résumés (elle n'est plus une classe à part entière dans la classification « d » car elle y est découpée en deux morceaux)	Comme déjà indiqué, il s'agit d'une spécificité française... et elle n'apparaît pas dans les classes c ou d.	Cette classe renvoie à la fois aux classes d6 et d7 (qui découpent ce thème en deux parties : <i>school</i> pour d6 et <i>college</i> pour d7).	Cette classe est en partie reliée à la classe d8, centrée « Santé ».

D'autres tableaux similaires France/USA pourraient être construits pour chacune des analyses présentées dans cet article. Ils renforceraient chacun le constat que l'on peut d'ores et déjà faire à partir de ce tableau 5 et des résultats des arbres des similitudes. Nous en retiendrons trois ici :

- (1) une différence majeure : la description systématique des méthodes dans les résumés USA,
- (2) des transversalités de thématiques : de la formation des maîtres, de la santé et des langues ainsi que des transformations biographiques,
- (3) l'existence de spécificités nationales comme l'analyse de l'activité et la didactique pro ou la VAE pour la France.

A cela s'ajoute une importante similitude qui apparaît en creux entre la situation de la France et des USA : le fait qu'il n'y ait ni dans un pays ni dans l'autre de marqueur évident de ce qui est de la FdA ou n'en est pas. En effet, en conduisant un tel travail, on ne peut qu'être pourchassé par le sentiment que l'on pourra toujours inclure de nouveaux mots clefs dans les requêtes pour aller chercher un peu plus loin d'autres travaux intéressants. Pour autant, cela n'invalide en rien le fait que ce travail est une nécessité pour quiconque se pose la question de l'état de l'art en matière de recherche en FdA.

5. Conclusion provisoire

Comme déjà indiqué, ce nouvel article consacré aux thèses s'intéressant à la FdA se fixait comme objectif d'aborder les objets de recherche et méthodes employées dans les thèses étatsuniennes, et ce au cours des 20 dernières années, c'est-à-dire 10 ans de plus que le corpus exploité dans VdR8. Dans le but de pouvoir les comparer le plus rigoureusement possible aux thèses françaises, nous avons

également travaillé avec deux corpus concentriques, d'abord un noyau de thèses entièrement dédiées à la FdA, puis un cercle élargi de thèses qui aborde de façon significative des travaux de recherche sur la FdA, sans pour autant que cette thématique soit explicitement affichée dans leur titre. Rappelons que ceux-ci sont accessibles en open access (comme tous les traitements et graphiques évoqués ici) via le Q/R code placé à la fin de cet article.

- difficultés rencontrées

Sans surprise, les difficultés rencontrées ont été globalement du même ordre que celles exprimées dans la conclusion de VdR8. Elles sont d'ordre linguistique, mais également d'ordres socio-politique et bibliographique. A ce dernier niveau, au-delà des obstacles rencontrés dans VdR 8, le fait que la base PQDT de Proquest relève depuis 2021 d'un abonnement très onéreux, sans possibilité de requêtes intelligentes, a rendu le corpus difficile d'accès et la recherche fastidieuse. Enfin, il nous a manqué de pouvoir appréhender l'observation de l'ensemble des résultats obtenus au prisme du regard de collègues étatsuniens.

- constats et apports relatifs à la comparaison France/USA

On peut souligner une différence majeure entre les résumés des thèses françaises et étatsuniennes : ces dernières mettent bien plus en avant les méthodologies utilisées. Il serait sans doute trop rapide d'en conclure que ces thèses sont plus solides que les thèses françaises d'un point de vue méthodologique. Il se peut tout simplement que les grandes lignes de la méthodologie soient exigées dans les résumés américains, contrairement à ce qui se passe en France.

Autre différence importante, si l'on considère les noyaux des thèses françaises et étatsuniennes, c'est la centration des premières sur la « *formation* » et des suivantes sur l'« *individu* ».

Concernant les thématiques à proprement parler, si l'on considère les résultats mis en avant dans la figure 17 et ceux de VdR7, on pourra observer que celles qui sont abordées communément par les thèses soutenues dans les deux pays relèvent principalement de l'échelle méso : santé, langues, formation des enseignants et transformations biographiques. Par ailleurs, la présence de la thématique de la formation en ligne est bien présente à 15% dans les thèses américaines (si l'on considère les titres et mots clés), alors que nous en avons observé l'absence dans les françaises (VdR7).

En termes d'évolution : Si l'on n'observe pas d'évolution notable des thématiques abordées dans les thèses françaises au cours des dix dernières années (VdR7), ça n'est pas le cas pour les thèses américaines : les thématiques « *andragogy, instructional design* » et « *colleges* » sont en croissance tandis que « *art, social et transformative* » et surtout la classe « *health* » sont en décroissance. Rappelons aussi le constat de la décroissance aux USA des méthodes quantitatives au profit des qualitatives.

En l'état, nous disposons donc de résultats intéressants et qui méritent encore d'être analysés plus avant. Ceci pourrait se faire en collaboration avec des universitaires experts de la FdA aux USA. Nous pourrions alors tirer pleinement parti des données recueillies et en faire état dans un VdR10 qui clôturerait ce travail.

Références bibliographiques

International Affairs Office (2020) *Structure of the U.S. Education System: Research Doctorate Degrees*. En ligne à

<https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww2.ed.gov%2Fabout%2Foffices%2Flist%2Fous%2Finternational%2Fusnei%2Fus%2Fdoctorate.doc&wdOrigin=BROWSELINK>

Las Vergnas, O. & Bury, P. (2020a). « Repérage des thèses intéressant les sciences de la formation des adultes : identification d'un noyau dur et d'un second cercle de thèses à consulter » (VdR6). *Savoirs*, 53, 89-110. Doi : 10.3917/savo.053.0089

Las Vergnas, O. & Bury, P. (2020b). « Analyse du corpus de référence des thèses françaises concernant la recherche sur la formation des adultes soutenues depuis 2010 : thématiques, rattachements et spécificités » (VdR7). *Savoirs*, 54, 83-107. Doi : 10.3917/savo.054.0083

Las Vergnas, O., Bury, P. & Jeunesse, C. (2022). "Recherches doctorales intéressant la formation des adultes : Repérage et première analyse de thèses soutenues aux USA entre 2010 et 2020 (VdR8)." *Savoirs*, 58.1, 67-99.

Ratinaud P. & Déjean S. (2009). IRaMuTeQ : implémentation de la méthode ALCESTE d'analyse de texte dans un logiciel libre. Modélisation Appliquée aux Sciences Humaines et Sociales (MASHS2009). Toulouse - Le Mirail.

Ratinaud P. & Marchand P. L'analyse de similitude appliquée aux corpus textuels : les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). 11èmes Journées internationales d'Analyse statistique des Données Textuelles, 2012, Liège, Belgique. 687-699. En ligne à <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Marchand,%20Pascal%20et%20al.%20-%20L'analyse%20de%20similitude%20appliquee%20aux%20corpus%20textuels.pdf>

Reinert M. (1987). Un logiciel d'analyse lexicale. *Cahiers analyse des données*, 11-4, 471-484. En ligne à http://www.numdam.org/item/CAD_1986__11_4_471_0

Vergès, P. et Bouriche, B. (2001). L'analyse des données par les graphes de similitude. *Sciences Humaines* (en ligne : <http://www.scienceshumaines.com/textesInedits/Bouriche.pdf>).

Ouverture des données

1) Toutes les figures (en version couleur) ainsi que tous les tableaux et fichiers créés pour ce chapitre ou simplement cités sont disponibles (au format tableur) en ligne en archives ouvertes via http://www.parhelie.org/publis/2022_VdR9 et le QR code ci-dessous.



2) Les requêtes décrites ont été effectuées au cours du mois d'août 2022 depuis une console connectée en local à <http://proquest.com> sur le réseau de la bibliothèque de recherche de la BNF. Toutes les extractions ont été mises à jour le 30 août 2022. Nous en profitons pour remercier ici toutes les personnes des bibliothèques de Paris-Nanterre, de Lille et de la BNF qui nous ont aidé à chercher un moyen d'avoir accès à la base PQDT.