



**HAL**  
open science

## N ° 3 - Travaux de recherche accessibles via l'archive ouverte de référence HAL (mars 2017)

Olivier Las Vergnas

### ► To cite this version:

Olivier Las Vergnas. N ° 3 - Travaux de recherche accessibles via l'archive ouverte de référence HAL (mars 2017). *Savoirs : Revue internationale de recherches en éducation et formation des adultes*, 2017, Les communautés d'apprentissage : apprendre ensemble, 43, pp.88-105. 10.3917/savo.043.0087 . hal-01625759

**HAL Id: hal-01625759**

**<https://hal.science/hal-01625759v1>**

Submitted on 28 Oct 2017

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

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



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

## **Rubrique « Vie de la Recherche en Formation des adultes » commune aux Revues *Savoirs* et *TransFormations***

**N°3 – Travaux de recherche accessibles via l'archive ouverte de référence HAL (mars 2017)**

***Olivier Las Vergnas, CIREL-Trigone (EA 4354) et CREF-AFA (EA 1589).***

### *Résumé :*

Cet article est le troisième de cette rubrique. Après deux investigations consacrées aux thèses, celui-ci se propose d'identifier et de décrire les publications déposées dans l'archive ouverte HAL qui est aujourd'hui considérée comme un entrepôt de référence pour les publications scientifiques francophones. Pour ce faire, deux méthodes de recherche de termes significatifs dans les résumés et listes de mots clés sont mises en œuvre et comparées. D'une part, un premier corpus (appelé « corpus générique ») est obtenu via une recherche sur la présence de l'un ou l'autre des termes génériques que sont « formation des adultes », « éducation des adultes », « formation tout au long de la vie » ou « éducation permanente ». D'autre part un autre corpus (appelé dans la suite « corpus issu du *Traité* ») est obtenu par une définition « extensive » : une batterie de requêtes teste la présence d'un ou plusieurs mots issus une liste de termes significatifs des divers aspects de la formation des adultes ; ces termes sont issus du sommaire d'un ouvrage pouvant être considérés comme un référentiel du champ (en l'occurrence le *Traité des sciences de la formation* de Carré et Caspar, troisième édition, 2011).

Le corpus générique va agréger 172 publications distinctes alors que celui issu du *Traité* va en regrouper 1182, soit huit fois plus, dont seulement 55 en commun. Malgré cette différence quantitative et ce faible recouvrement, ces deux corpus se révéleront présenter des spectres de thématiques qualitativement assez semblables, caractérisés par un triple noyau : « technologie et pédagogie », « certification et validation » et « socio-économie la formation » ; on observera aussi une meilleure démarcation à l'intérieur du corpus issu du *Traité* d'une thématique « ergonomie et didactique pro ».

La première conclusion est que, dans HAL comme dans le Sudoc (cf n° précédents), la plupart des publications (ici 85 %) sont indexées trop finement pour être retrouvées par une seule requête générale ; une autre conclusion est que -même si les différentes acceptions de ce champ fluctuent largement- celui-ci semble présenter dans HAL ce triple noyau relativement stable.

*Mots clés :* publications de recherche, bibliométrie, formation des adultes, analyse lexicale.

*Note de l'auteur :* les données rapportées ici correspondant à des requêtes effectuées dans HAL-SHS fin février 2017.

## **1. Position par rapport au contexte et au travail lancé dans cette rubrique**

### 1.1. Élargissement de l'analyse au-delà des thèses

Cet article est le troisième de la nouvelle formule<sup>1</sup> de la rubrique « vie de la recherche en formation des adultes » ; il s'inscrit donc dans la volonté globale de proposer un état dynamique des lieux de la production de connaissances dans ce champ, au moins dans l'univers francophone pour le moment. Les deux premiers articles ayant été fondés sur l'analyse des thèses répertoriées dans la base de données du Service Universitaire de Documentation (Sudoc), l'analyse est ici poursuivie en élargissant la base de documents explorés. Le but est en particulier d'inclure des articles, ouvrages et parties d'ouvrages et autres communications scientifiques plus collectifs et plus courts que des thèses.

### 1.2. Montée en puissance des archives ouvertes, des moteurs spécialisés et des réseaux sociaux

Dans cet objectif de diversification, la base du Sudoc précédemment utilisée se révèle peu pertinente car les articles de revues scientifiques SHS n'y sont que très peu indexés. Il faut donc s'appuyer sur une autre source et plusieurs solutions sont envisageables, d'autant que le contexte actuel voit une triple montée en puissance d'outils complémentaires : (1) celle des moteurs de recherche spécialisés dans le moissonnage de publications (généralistes comme Google Scholar ou spécialisés comme Eric ou Isidore) ; (2) celle des réseaux sociaux spécifiques (comme Academia ou Research Gate) (3) celle des entrepôts d'archives ouvertes (comme HAL) dont le déploiement et l'usage systématique sont maintenant encouragés. En croisant plusieurs paramètres (compatibilité avec des sélections francophones, volume de données disponibles, reproductivité des requêtes, maîtrise du périmètre, contrôle de fiabilité des métadonnées), c'est l'exploration de l'archive ouverte HAL (Bouchard, 2016) et plus particulièrement du portail HAL-SHS qui a été retenue pour le numéro 3 de cette rubrique.

### 1.3. L'archive ouverte « Hyper archive en ligne » (HAL)

Ce système pluridisciplinaire d'archive et de référencement porté par le centre pour la communication scientifique directe du CNRS est « destiné au dépôt et à la diffusion d'articles scientifiques de niveau recherche, publiés ou non, et de thèses, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés » (cf <https://halshs.archives-ouvertes.fr/> ). Il constitue le premier « entrepôt » pluridisciplinaire francophone en *open access* pour le dépôt volontaire par les chercheurs et leurs institutions de document, avec plus de 58532 documents déposés en 2016, dont 20506 articles publiés dans des revues, 13966 communications dans des colloques et 7943 thèses. De plus, HAL ne donne pas seulement accès aux textes intégraux d'articles, mais référence plus de 3 fois plus de publications : les requêtes HAL accèdent au total à plus 1 230000 références dont plus de 430000 documents intégraux déposés (Magron, 2017).

Bien sûr, comme pour c'est le cas pour toute base, les textes déposés ou référencés dans HAL SHS ne peuvent être vu comme représentatifs de toutes les publications de recherche : HAL est sans doute plus utilisé par les chercheurs et/ou institutions les plus soucieux de leur accessibilité et pérennité en ligne. Les résultats qui suivent doivent donc se regarder comme une photographie partielle favorisant les recherches sur la formation des adultes que de tels chercheurs ou institutions ont rendu accessibles (ou repérables) via les archives ouvertes.

Il se trouve que ce type d'archives ouvertes constitue un des fers de lance des politiques d'*open access* dont le développement est souhaité aujourd'hui par les pouvoirs publics ; en suivre la montée en puissance par rapport à d'autres outils va devenir d'un grand intérêt bibliométrique (Bouchard, 2016), d'autant que les sous-portails institutionnels des laboratoires sont en train de se développer et de s'organiser. Alors que dans la période des débuts de HAL les dépôts, en particulier en SHS, ne résultaient que de choix individuels de chercheurs ceux-ci deviennent maintenant facilités voire

---

<sup>1</sup> En accord entre les rédactions des deux revues, cette rubrique est commune à la revue « Savoirs » et à la revue « TransFormations ».

orchestrés par leurs laboratoires et établissements, dans un élargissement similaire à celui de la mobilisation de la VAE par les DRH et non pas seulement par les individus. L'ambition ici est donc non seulement de fournir un instantané, mais aussi de donner un premier point de repère, voire une méthodologie pour des études longitudinales ultérieures.

## **2. Présence des thématiques liées à la « formation des adultes » et double stratégie**

Comme l'ont montré les deux numéros précédents, les travaux sur la formation des adultes sont fréquemment signalés ou indexés par des termes très spécifiques ; en conséquence, la recherche des publications doit se faire non seulement par des requêtes génériques cherchant dans le résumé ou les mots clefs la présence d'expressions larges comme « formation des adultes », « formation permanente » ou « formation tout au long de la vie » mais aussi via des séries de requêtes plus pointues sur des termes précis, comme par exemple « validation des acquis » ou « apprentissage en situation de travail ».

Formuler de telles requêtes focalisées revient à déterminer par extension une définition de la formation des adultes. Éviter une approche autoréférente demande donc de s'appuyer soit sur une liste d'équipes ou d'auteurs, soit sur une liste préétablie de thématiques pouvant faire office de référentiel de définition. C'est ce qui a été fait ici, où les requêtes d'interrogation ont été construites à partir du sommaire d'un ouvrage de référence du champ, en l'occurrence le « *Traité des sciences de la formation* » (Carré et Caspar, 2011, troisième édition), désigné dans la suite par « *Traité* »

Ainsi, nous avons cherché à constituer deux sous-corpus de publications déjà référencées dans HAL en fonction de termes se trouvant mentionnés soit dans leur résumé, soit dans leurs mots clefs : (1) un premier sous-corpus (appelé dans la suite « corpus générique » regroupant celles qui mentionnent l'un au moins des termes génériques « formation des adultes, éducation des adultes, formation permanente, formation tout au long de la vie »; (2) un second avec celles mentionnant au moins l'un des termes correspondant au sommaire du *Traité*, toujours soit dans leur titre ou mots clefs : nous avons utilisé les mots signifiants des titres de chapitres pour constituer 17 paires de requêtes (une pour chercher dans les mots clefs et l'autre dans le résumé) et nous avons d'ailleurs ajouté une paire de requêtes correspondant au terme « référentiel » et une autre à l'expression « gestion prévisionnelle des emplois et des compétences », tous les deux caractéristiques de deux dimensions non explicitement formulées dans les titres des chapitres du *Traité*. Bien sûr, cette méthode conduit à des doublons, à savoir des publications qui sont repérées par plusieurs requêtes différentes (entre 2 à 6 fois au maximum) et que nous allons choisir de garder pour le moment dans le corpus.

## **3. Au total, 1806 références dont 1299 publications réellement distinctes**

Au total, le regroupement de toutes ces requêtes, génériques ou non, fournit un ensemble initial constitué de 2135 références, ce après élimination de 107 qui n'avaient ni résumé, ni titre et d'une pittoresque série de 29 images représentant des vues d'une usine de chaudronnerie dont la notice mentionnait « compétence » et « formation »).

### **3.1. Filtrage supplémentaire concernant la formation des enseignants**

Une analyse lexicale de vérification de cet ensemble initial montre qu'il y subsiste de nombreuses références à la formation initiale, correspondant par exemple à des publications mélangeant développement professionnel des enseignants et travail des élèves en milieu scolaire ; nous les éliminons du corpus par une nouvelle requête filtre (pas « enseignant\* ») ce qui en écarte encore 328 (dont 32 avaient été fournies par la requête générique). Finalement, le corpus retenu est réduit à 1806 citations dont 1299 correspondent à des documents différents. Les tableaux 1 et 2 donnent, requête par requête le nombre de publications repérées dans HAL. Ils en décrivent 1797 sur les 1806 trouvées car ils ont été allégés des deux dernières colonnes, qui contenaient les 9 manquantes (correspondant aux types - très rares dans HAL- que sont « images » et « lectures » d'ouvrage, ainsi qu'à un document non classé). En fournissent un spectre de thématiques ils constituent un premier résultat de ce travail.

	COMM	ARTICLE	THESE	Ch d'OUV.	AUTRE	UNDEFINED	OUV	Dr d'OUV	REPORT	HDR	Total général
Chapitre 13 Compétence et formation (Résumé)	117	91	56	43	19	21	12	9	11	4	383
Chapitre 13 Compétence et formation (Mots clefs)	43	41	17	15	9	1	5	3	3		137
<b>GENÉRIQUE « Formation des adultes » et assimilé (Résumé)</b>	<b>20</b>	<b>30</b>	<b>9</b>	<b>12</b>	<b>10</b>	<b>7</b>	<b>5</b>	<b>5</b>	<b>1</b>	<b>1</b>	<b>100</b>
Chapitre 20 Professionnalisation et formation (Résumé)	30	22	12	13	1	6	8	2	3	1	98
<b>GENÉRIQUE « Formation des adultes » et assimilé (Mots clefs)</b>	<b>22</b>	<b>25</b>	<b>6</b>	<b>23</b>	<b>6</b>	<b>3</b>	<b>3</b>	<b>6</b>	<b>2</b>	<b>2</b>	<b>98</b>
Chapitre 18 Autoformation (Mots clefs)	19	25	8	26	1		2	3	2		86
Chapitre 15 Motivation et formation (Résumé)	21	22	21	5	3	2	3	2	2	1	82
Chapitre 18 Autoformation (Résumé)	24	23	5	16	1	2	1	1	3	1	77
Chapitre 14 Mémoire et formation (Résumé)	11	11	23	5	1	1	3	3		7	65
Chapitre 20 Professionnalisation et formation (Mots clefs)	11	16	6	15	2			4	1		55
Chapitre 7 Management des orga. et formation (Résumé)	8	10	25		1	2	2		2	2	52
Chapitre 29 VAE (Mots clefs)	10	13	1	14	5	2	3				48
Chapitre 29 VAE (Résumé)	5	17	3	7	4	2	4	2	3		47
Chapitre 22 Développement prof. et formation (Résumé)	22	11	2	4						1	40
Chapitre 23 Didactique pro. et formation (Mots clefs)	15	15	5	1		1				1	38
Chapitre 6 Gestion des RH et formation (Résumé)	13	7	4	3	2	1	3				33

Tableau 1 : Nombre de publications dans HAL selon les thématiques : requêtes correspondant à plus de 30 publications. Les lignes commencent par le N° du chapitre ; la mention qui suit indique si le terme clef est recherché dans le résumé ou les mots clefs ; suit le type (article, ouvrage individuel, chapitre d'ouvrage, direction d'ouvrage, thèse, HDR, poster). Les requêtes utilisées ([api.archives-ouvertes.fr/search/](http://api.archives-ouvertes.fr/search/)) bloquent de plus celles ne contenant pas « adultes » et « formation ».

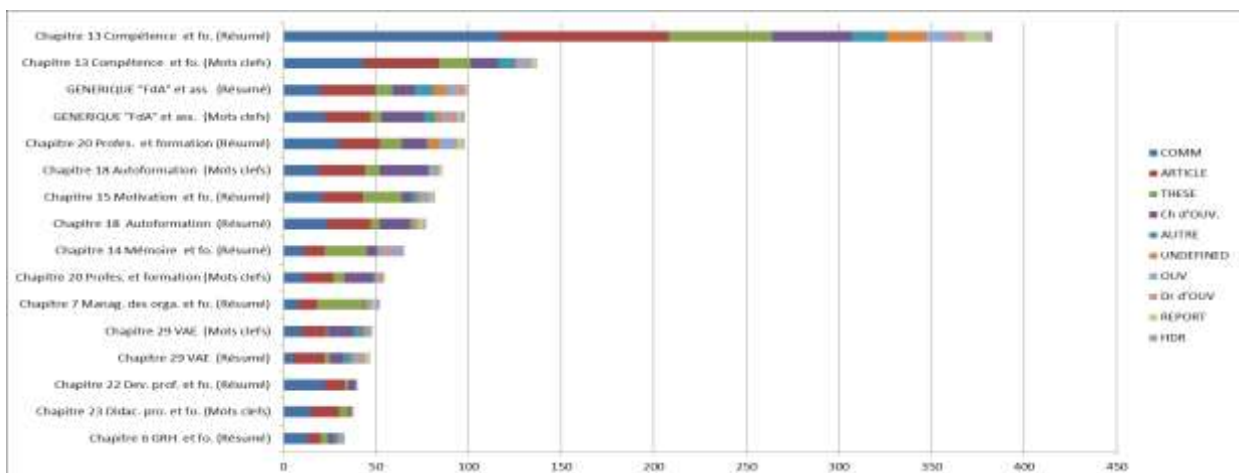


Figure 1 : Nombre de publications dans HAL selon les thématiques : Requêtes correspondant à plus de 30 publications.

En terme de diversité des types, on voit (figure 2) que ces requêtes fournissent au total près de deux fois plus de « communications dans un colloque » (27%) et d'articles scientifiques (26%) que de thèses (14%) ou de chapitres (14%).

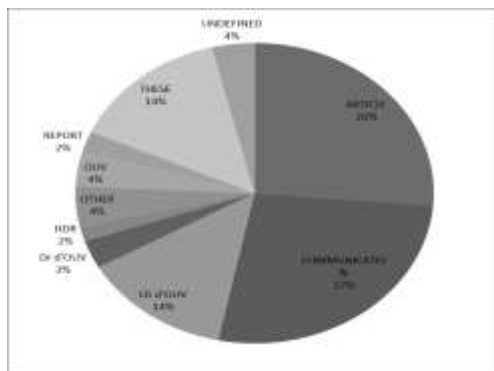


Figure 2 : Répartition des publications par type dans HAL (sur le corpus total de 1807 références)

On voit également que le cumul des thèses trouvées (toutes requêtes confondues) est de 260 (et celui des HDR de 29), ce qui est comparable au nombre total de thèses identifiées (275) dans le Sudoc via des requêtes génériques lors du précédent article de cette rubrique<sup>2</sup>.

	COMM	ARTICLE	THESE	Ch d'OUV.	AUTRE	UNDEFINED	OUV	Dr d'OUV	REPORT	HDR	Total général
Chapitre 25 IngePedago et formation (Mots clefs)	6	8	4	3		1	1	4	1	1	29
Chapitre 23 DidacPro et formation (Résumé)	8	15	4			1				1	29
Chapitre 22 DevelopPro et formation (Mots clefs)	11	11	1	2				1	1		27
Chapitre 30 FormateurAdult (Mots clefs)	4	7	1	8		1	1	2			24
Chapitre 12 Viellissement et formation (Résumé)	4	4	2	3	1	3		2	1	1	21
Chapitre 25 IngePedago et formation (Résumé)	4	7	4	1		2		2			20
Chapitre 21 Politique et formation (Mots clefs)	6	5	1	3	2			1	1		19
Chapitre 30 FormateurAdult (Résumé)	9	1	4	2		1	1				18
Chapitre 8 ErgonoDvForm (Résumé)	5	1	3			3	1		1	3	17
Chapitre 15 Motivation et formation (Mots clefs)	3	1	5	2		2		1			14
Chapitre 14 Mémoire et formation (Mots clefs)	3	4	3	1	1					1	13
Chapitre 21 Politique et formation (Résumé)	2	1	4	2	1		1	1			12
Partie II Intro : Apprenance_ (Résumé)	3	2	4								9
Chapitre 10 Diversité et formation (Résumé)	1	3			1		2	1			8
Chapitre 9 EuropForm (Résumé)	1	2		3	2						8
Chapitre 2 EcoForm (Mots clefs)	2	3		1					1	1	8
Addendum : Referentiels (Résumé)	1	2	2						1		6
Chapitre 41 BilanComp (Résumé)	1	2	2						1		6
Chapitre 5 MarcheForm (Résumé)	1	3			1						5
Chapitre 16 ConflitSoc et formation (Mots clefs)	3	1		1							5
Chapitre 6 GestionRH_ (Mots clefs)	2	1			1		1				5
Chapitre 7 ManagOrga (Mots clefs)				3		1	1				5
Chapitre 16 ConflitSoc et formation (Résumé)	3	1		1							5
Chapitre 41 BilanComp (Mots clefs)	2	1	1								4
Chapitre 11 VieAdulte et formation (Résumé)	1	1	2								4
Addendum : Referentiels (Mots clefs)	2	1	1								4
Chapitre 3 SocioForm_ (Mots clefs)		1	2								3
Chapitre 0DetermEnvir (Résumé)			1	1			1				3
Chapitre 5 MarcheForm_ (Mots clefs)	1	1				1					3
Chapitre 8 ErgonoDvForm (Mots clefs)	2	1									3
Chapitre 24 AppSituTrav (Mots clefs)		1	1	1							3
Chapitre 1 PostSco (Résumé)			1	1			1				3
Chapitre 12 Viellissement (Mots clefs)	1					1					2
Chapitre 28 OrientProf (Résumé)	1				1						2
Chapitre 4 DroitForm (Résumé)			1		1						2
Chapitre 4 DroitForm (Mots clefs)		1		1							2
Chapitre 24 AppSituTrav (Résumé)			2								2
Chapitre 28 OrientProf (Mots clefs)			1								1
Chapitre 19 PsychoPeda et formation (Mots clefs)						1					1
Partie III (intro) ConduiteFor (Résumé)		1									1
Chapitre 1 PostSco (Mots clefs)							1				1
Chapitre 3 SocioForm (Résumé)				1							1
<b>Total général (y compris des lignes &gt;30 et génériques)</b>	<b>484</b>	<b>473</b>	<b>260</b>	<b>243</b>	<b>77</b>	<b>69</b>	<b>66</b>	<b>55</b>	<b>41</b>	<b>29</b>	<b>1797</b>

Tableau 2 : Nombre de publications dans HAL selon les thématiques ; requêtes inférieures à 30 publications.

<sup>2</sup> Si nous restons dans l'hypothèse émise alors -à savoir que ce nombre est sous-estimé de moitié et qu'il y en aurait en fait 550 dans le Sudoc- on peut imaginer que ce nombre de 260 s'explique par le fait que seule une petite moitié de ces 550 thèses référencées dans le Sudoc ont été entrées dans HAL ce qui revient à dire que seul un sur deux de ces 550 anciens doctorants a aujourd'hui entré sa thèse dans HAL.

### 3.2. un huitième seulement des publications repérées par des termes génériques

Comme dans le cas des thèses du Sudoc les publications mentionnant les termes génériques que sont « formation des adultes », « éducation des adultes », « formation tout au long de la vie », « éducation permanente » dans leur résumé ou mots clefs ne représentent qu'une petite fraction de celles repérées par les termes des chapitres du *Traité* et donc a fortiori de celles qu'une analyse heuristique humaine pourrait considérer comme concernant ce sujet : les lignes 3 et 5 du tableau 1 indiquent en effet que l'on ne repère dans HAL que 100 documents pour lesquels ces termes génériques sont présents dans le résumé et 98 dans les mots-clefs.

De plus, au sein de ce total de 100+98 références « génériques », il s'en trouve 36 qui se révèlent en doublon entre ces deux requêtes (en raison de la présence simultanée de termes génériques dans le résumé et les mots clefs) : cet ensemble ne rassemble ainsi que 172 publications distinctes. Or, à titre d'exemple dans le même temps, les publications contenant « compétences » et « formation et adulte » sont beaucoup plus nombreuses : 383 dans leur résumé et 137 dans leurs mots-clefs.

### 3.3. 55 doublons entre les publications repérées par les termes génériques ou celles par des spécifiques

En réalité, après vérification on se rend compte que, comme l'illustre la figure 2, les deux sous corpus (générique et lié au *Traité*) partagent 55 références en commun, qui contiennent donc au moins un terme générique et un terme caractéristique d'au moins un chapitre.

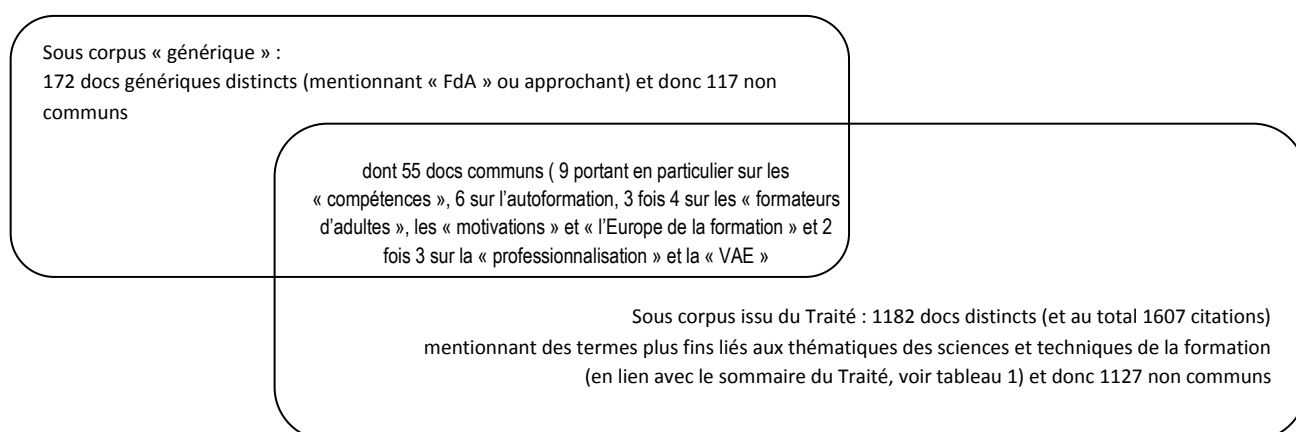


Figure 3 : Répartitions des 1299 documents retenus dans HAL-SHS concernant les thématiques de la « formation des adultes ».

Comparer les efficacités de repérage entre requêtes génériques et requêtes sur les mots du sommaire du *Traité*, revient à mettre en perspective les 1299 publications différentes avec les 55 communes (donc trouvées simultanément par les deux méthodes) ou avec les 172-55 = 117 trouvées uniquement par les requêtes génériques. De fait, sur 1182 publications entrant dans le champ au sens strict des mots du *Traité*, seules 4.2% (55 communes sur 1182, dont 9 portant en particulier sur les « compétences », 6 sur l'autoformation, 3 fois 4 sur les « formateurs d'adultes », les « motivations » et « l'Europe de la formation » et 2 fois 3 sur la « professionnalisation » et la « VAE ») sont repérés aussi par les requêtes génériques. De plus, un volume équivalent à près de 10% de ces 1182 (117 non communes par rapport à 1182) n'est trouvé par aucune des 17 paires de requêtes sur des mots du *Traité*<sup>3</sup>. Voilà de quoi convaincre que cette question de l'indexation est cruciale.

<sup>3</sup> Concrètement, cela signifie que 95% des publications concernant la formation des adultes au sens strict du *Traité* ne sont bibliométriquement identifiables que par des mots spécifiques pointus pris dans les titres des chapitres et qu'inversement l'équivalent de 10% parlant autrement de la formation des adultes ne le sont pas.

#### 4. Analyses lexicales des deux sous corpus

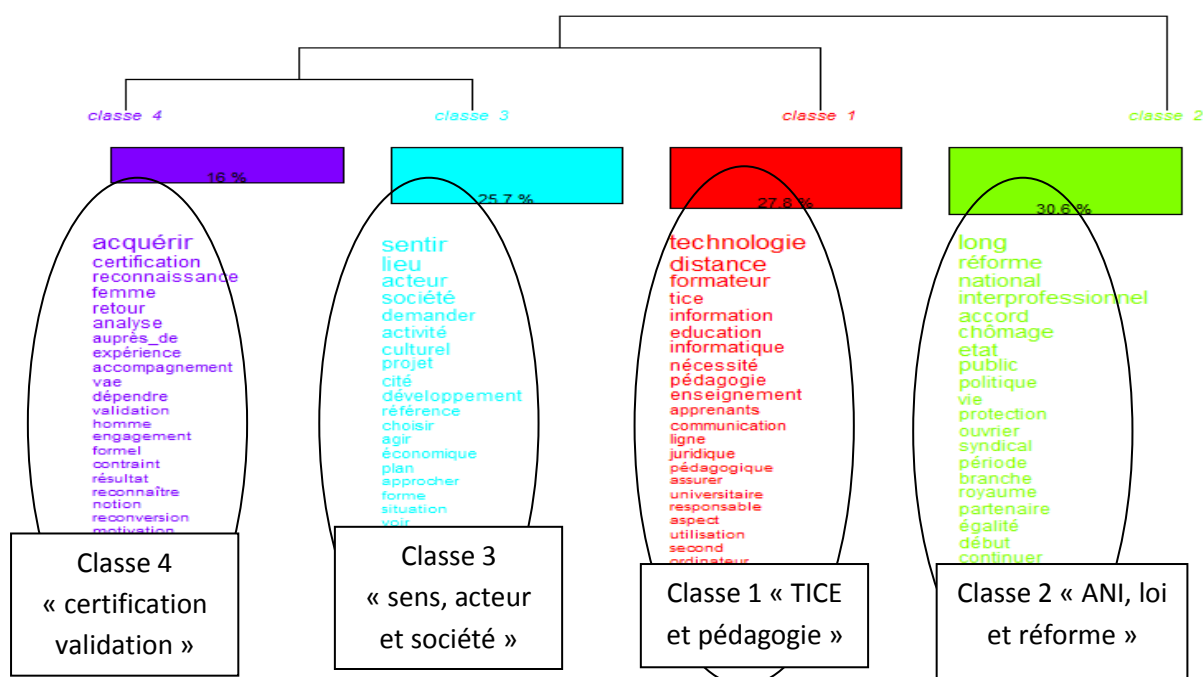
Pour chacun de ces deux sous-corpus, nous avons conduit un travail d'analyse lexicale des résumés et titres. Comme lors les deux premiers numéros de cette rubrique, le logiciel Iramuteq et les classifications s'appuyant sur les algorithmes de Reinert (Reinert, 1987 ; Ratinaud et al. 2009) ont été utilisés. L'idée est de comparer les classes de lexiques identifiées comme correspondant à l'un et à l'autre des deux sous-ensembles de publications afin de cartographier les thématiques de recherche leur correspondant.

Pour cela leurs caractéristiques et leurs points communs mais aussi leurs spécificités selon les requêtes utilisées doivent pouvoir être explorés : dans ce but, les citations en double entre les différents chapitres n'ont pas été supprimées : si une même publication HAL est repérée à la fois comme liée à deux (voire plus) chapitres, plutôt que de l'attribuer arbitrairement à l'un ou l'autre (et de l'enlever pour l'autre –ou les autres-), elle est maintenue dans chacun de façon à obtenir une cartographie non tronquée des recherches liées à chaque chapitre.

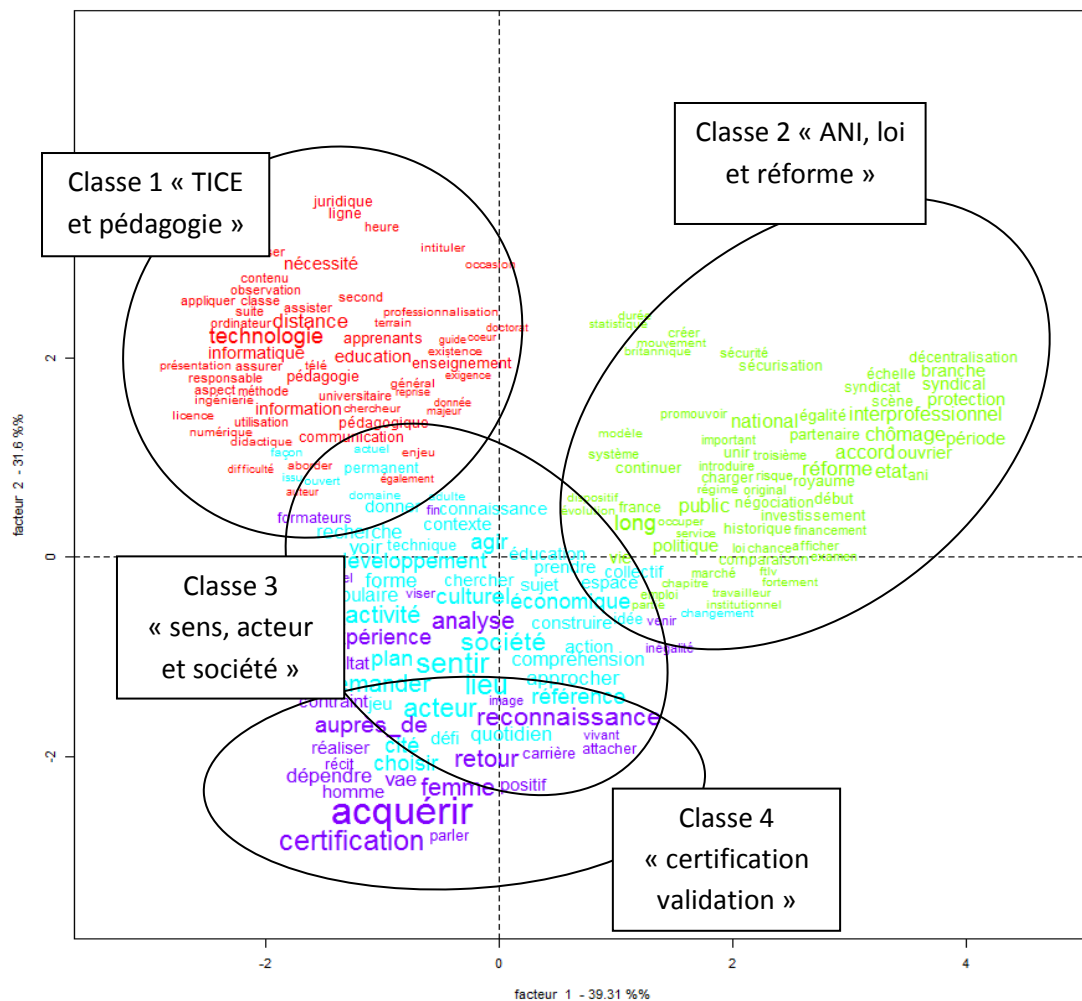
##### 4.1. Sous-corpus repéré par les termes génériques « formation des adultes » et assimilés

Une première analyse a été effectuée sur le corpus des 172 publications du corpus générique. Elle montre que les titres, résumés et mots clefs de ces 172 publications s'organisent selon des univers lexicaux liés à quatre grandes familles de préoccupations ou de centres d'intérêt qui sont présentées en figure 2 et 3.

On y observe un lexique (classe 4) lié à l'acquisition, la validation ou la certification, ainsi qu'un autre lexique proche (classe 3), lié aux lieux culturels, aux acteurs dans la société ; deux autres groupes lexicaux plus éloignés correspondent respectivement à (classe 1) aux TICE (technologie, formation à distance) et à la pédagogie et (classe 2) aux aspects socio-économiques comme la réforme de la formation ou l'accord national interprofessionnel.







Figures 2 et 3 : Classification des lexiques utilisés dans les 172 publications « génériques » i.e. mentionnant « formation des adultes » et approchant et déposés dans HAL-SHS (dendrogramme et plan 1,2 d'une AFC sur les classes –logiciel Iramuteq).

#### 4.3. Sous-corpus des publications repérées uniquement via les titres des chapitres du *Traité*

Nous avons parallèlement analysé l'autre sous-corpus des 1182 publications HAL, uniquement trouvées via les requêtes sur les titres de chapitres du *Traité* ; comme indiqué plus haut nous avons pris soin d'y laisser les différentes occurrences des références aux mêmes publications, lorsque celles-ci se révèlent apparaître en double ou plus, c'est-à-dire dans le cas où ces publications mentionnent plusieurs termes différents qui les renvoient à deux ou plusieurs chapitres distincts.

Les figures 4, 5 et 6 donnent les résultats d'une classification en 5 classes de ce sous-corpus. Celle-ci présente une structure en partie similaire à celle du sous-corpus générique. On y reconnaît d'abord une classe 5 similaire à la classe « TICE et pédagogie » et une classe 3 qui reprend l'essentiel des termes de la classe « certification et validation » en l'associant aux professions de santé ; on observe aussi une déformation de la classe socio-économique qui s'est déchirée en deux entre une partie centrale « emploi, marché, économie, politique, entreprise » sur laquelle s'est greffée une excroissance spécifique renvoyant des recherches sur l'histoire sociale. La figure 6 qui peut se superposer à la figure 5 montre la façon dont ces classes recourent sans surprise les thématiques du sommaire du *Traité*.

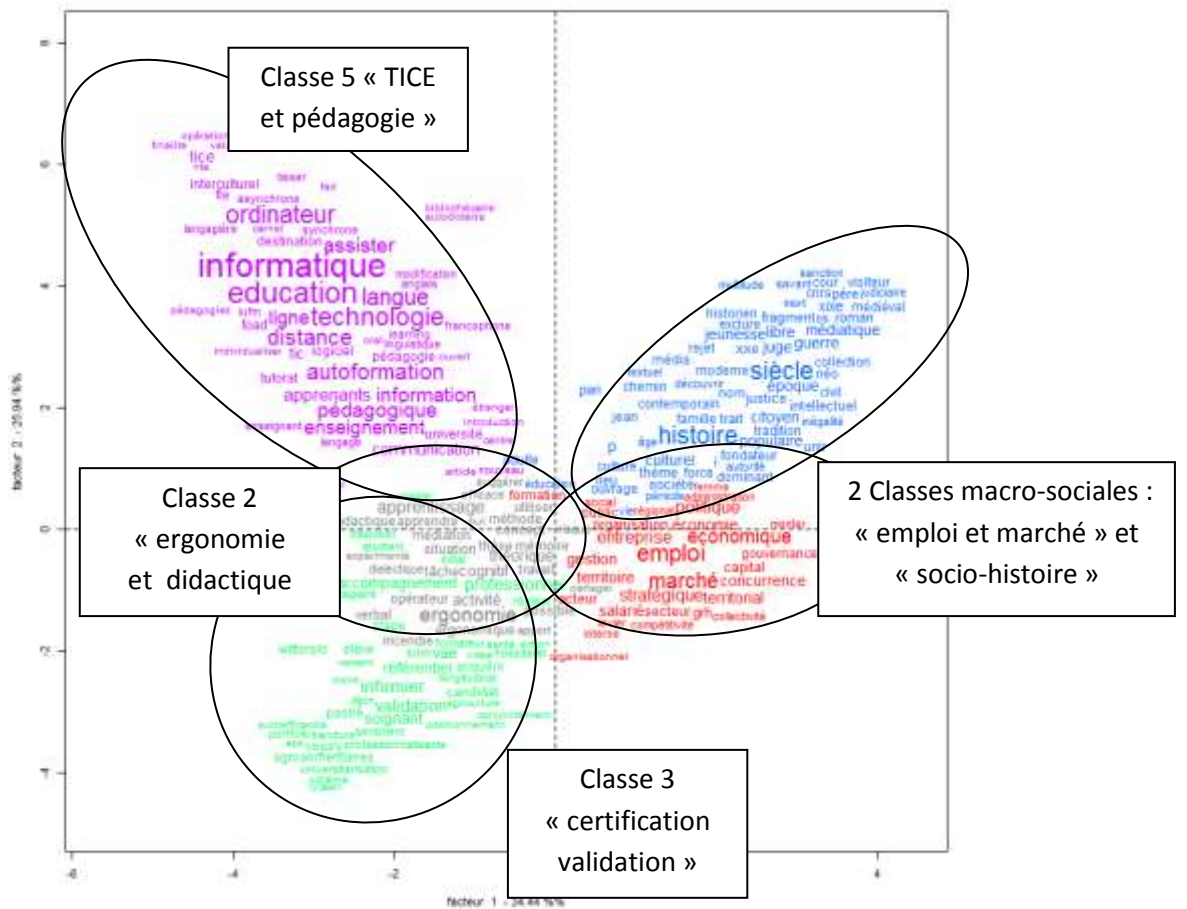
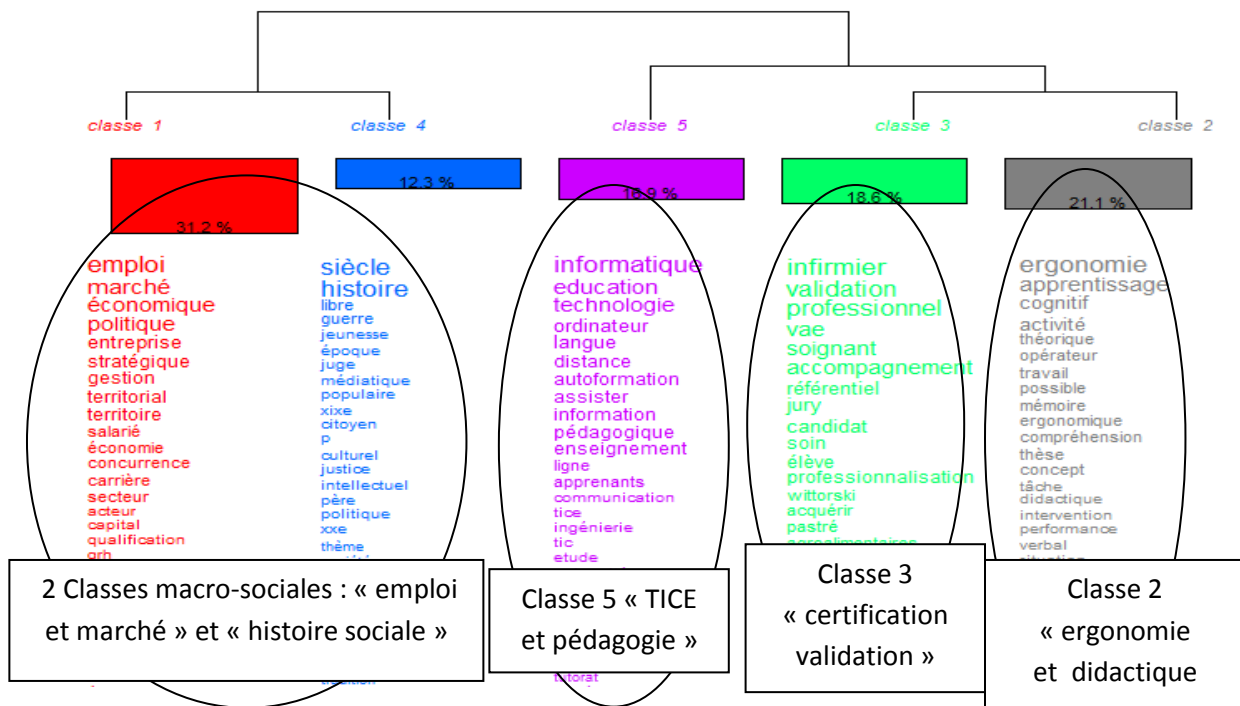


Figure 4 et 5 : Classification des lexiques utilisés dans les 1182 publications (1607 occurrences avec les doublons) mentionnant des mots signifiants des titres de chapitres du *Traité* et déposés dans HAL-SHS (logiciel Iramuteq).

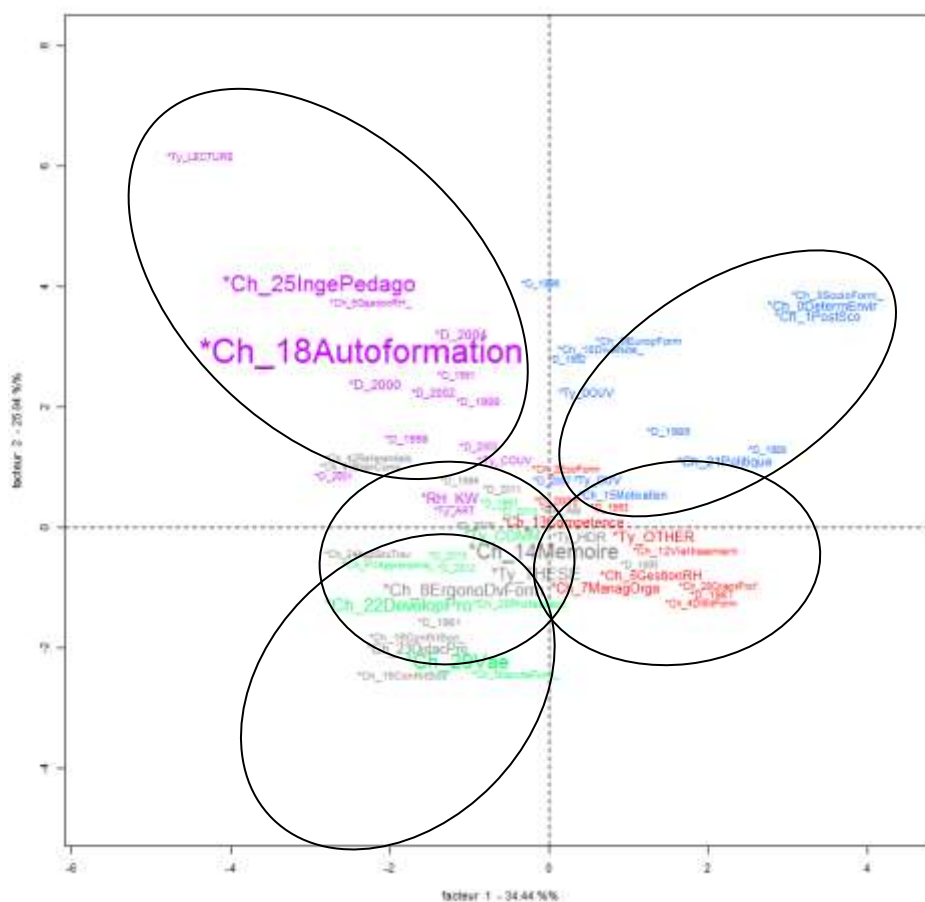


Figure 6 : Positionnement sur le plan factoriel de la figure 5 des termes utilisés dans les 1182 publications (1607 occurrences avec les doublons) en lien avec les chapitres du *Traité* (logiciel Iramuteq). Cette image est faite pour se superposer à celle de la figure 5 afin de voir la correspondance entre les lexiques et les thématiques du *Traité*.

## **5. Apports de l'analyse comparative des deux sous corpus**

Malgré cette différence quantitative et leur faible recouvrement, ces deux sous-corpus se révèlent présenter des spectres de thématiques qualitativement assez semblables, caractérisés par un triple noyau « technologie et pédagogie », « certification et validation » (relié plus spécialement dans le corpus issu du *Traité* aux professions de santé) et « socio-économie la formation » (qui se voit renforcé le corpus issu du *Traité* d'une excroissance spécifique à l'histoire sociale...); on observe aussi à l'intérieur du corpus *Traité* la démarcation d'une thématique « ergonomie et didactique pro ».

## **6. Première comparaison avec les corpus de thèses des deux numéros précédents et conclusion**

La première conclusion est que, dans HAL comme dans le Sudoc (cf n° précédents), la plupart des publications (ici 85 %) sont indexées trop finement pour être retrouvées par une seule requête générale; une autre conclusion est que -même si les différentes acceptions de champ fluctuent largement- celui-ci semble présenter dans HAL un triple noyau central relativement stable que conserve le petit échantillon qu'en représente le sous-corpus générique.

Bien sûr, ceci doit être comparé avec les travaux antérieurs conduits sur les thèses. Les lecteurs des numéros précédents vérifieront les différences de ce corpus avec celui des seules thèses indexées génériquement « éducation des adultes » en 2014/15: les classes liées à la formation continue des enseignants et aux apprentissages linguistiques ont disparu fort logiquement en raison du filtrage ajouté pour les éliminer; de plus une classe « économie – marché emploi » est apparue et la classe « sujet, identité, social » s'est transformée en « certification validation ». Ces constat suggère que la

différence principale des publications HAL (et sans doute aussi ou plutôt de leur indexation) par rapport aux thèses du Sudoc est le fait que l'« éducation des adultes » n'est plus seulement vue comme une reprise d'étude au sens du DAEU, ce qui semblait a contrario implicite dans l'indexation du Sudoc/Rameau. Pour aller plus loin dans ce sens, la comparaison plus fine de ces deux populations des Thèses du Suoc et des documents déposés dans HAL sera détaillée dans le prochain article de cette rubrique.

### Références bibliographiques

Bouchard A. (2016). *Éléments de comparaison archives ouvertes et réseaux sociaux*. Paris : URFIST <http://urfist.enc.sorbonne.fr/ressources/veille-sur-les-outils/diffuser-ses-travaux-sur-les-reseaux-sociaux-academiques-ou-sur-des>

Carré P. et Caspar P. (2011). *Traité des sciences et techniques de la formation*. (Troisième édition). Paris : Dunod.

Magron A. (2017). *HAL en 2016. Une activité soutenue*. CDSO Paris : CNRS <https://www.ccsd.cnrs.fr/fr/2017/01/hal-en-2016-une-activite-soutenue/>

Ratinaud P. et 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. Voir à [http://reperer.no-ip.org/Members/pratinaud/mes-documents/articles-et-presentations/presentation\\_mashs2009.pdf](http://reperer.no-ip.org/Members/pratinaud/mes-documents/articles-et-presentations/presentation_mashs2009.pdf)

Reinert M. (1987). Un logiciel d'analyse lexicale. *Cahiers analyse des données*, 11-4, 471-484. En ligne à [http://www.numdam.org/numdam-bin/fitem?id=CAD\\_1986\\_\\_11\\_4\\_471\\_0](http://www.numdam.org/numdam-bin/fitem?id=CAD_1986__11_4_471_0)