

Essai de classement des articles de revues d'auteurs du Cirad selon les 6 axes stratégiques

Présentation de l'étude menée par la DIST – 10/07/2008

Marie-Dominique Lafond, Marie-Claude Deboin, Catherine Gérard., Alain Glarmet, Annie Marti

1. Objectif général et contexte
2. Méthodologie
3. Résultats

1. Objectif général et contexte

Objectif et contexte

Cette étude **expérimentale** est réalisée par la DIST pour la DRS.

Trois objectifs

1. Étudier comment **classer** les publications du Cirad selon les 6 axes stratégiques à partir du référentiel institutionnel Agritrop.
En se limitant aux articles de périodiques 2007-2008.
2. Mémoriser cette information dans Agritrop pour qu'elle devienne
 - **pérenne**,
 - **accessible et réutilisable par tous**.
3. Rechercher comment automatiser totalement ou partiellement le classement et la saisie.

Objectif et contexte

Une étude réalisée par la
Délégation à l'information scientifique et technique (Dist)

Les participants à ce projet :

Marie-Dominique Lafond
(coordination)

Marie-Claude Deboin
Catherine Gérard
Alain Glarmet
Annie Marti

Chaque participant a classé l'ensemble du corpus analysé selon les 6 axes.

Ces professionnels de l'IST ont l'habitude d'indexer des documents scientifiques et techniques, en utilisant des référentiels de type plan de catégorisation et thesaurus.

2. Méthodologie

Méthodologie : abrégé

1. Caractérisation des axes : documents, mot clés

Les documents de références utilisés pour ce classement sont :

- Cirad - La vision stratégique 2008-2011 : recentrer les thématiques, renouveler les pratiques. - Montpellier : CIRAD, 2008. - 47 p.
- Le site internet du CIRAD : rubrique axes prioritaires
- Le site intranet de la DRS : les informations sur les axes et la liste des unités de recherche (au 30/06/2008)
- L'affiliation des auteurs au 29/04/2008 pour le classement des articles dans les UR
- Le thésaurus Agrovoc et le plan de catégorisation Agris/Caris (FAO)

Le groupe de travail a caractérisé chaque axe avec 8 mots clés au maximum :

- d'une part en langage naturel,
- d'autre part avec les descripteurs du thésaurus Agrovoc (utilisé pour la base Agritrop).

Méthodologie : abrégé

2. Constitution du corpus

Extraction d'Agritrop le 31/05/2008 des 800 articles de revues 2007 et 2008 ayant au moins un auteur avec une affiliation Cirad.

607 articles (2007) et 193 articles (2008)

3. Traitement des données

- **Test du classement** avec un premier lot de 94 publications. Analyse collective des résultats et mise au point de la méthode.
- **Classement** des 706 autres publications. Analyse des résultats et conclusions.
- Mise en place d'un **itinéraire technique** pour la saisie du « code axe stratégique » dans Agritrop pour ce corpus. Saisie du code dans Agritrop.
- **Relecture** de l'ensemble des articles regroupés par axe pour vérifier et homogénéiser le classement.

4. Analyse des résultats

- Répartition des articles par axe en nombre et en proportion.
- Répartition des articles suivant la **notoriété des revues** (facteur d'impact, comité de lecture) pour l'ensemble du corpus et pour chaque axe.
- Étude de la faisabilité d'un **classement automatique**
 - Répartition du nombre d'articles par **code Agris** et par axe
 - Pour **chaque UR** :
 - répartition du nombre d'articles par axe
 - identification d'un axe principal avec un seuil minimal de 70% d'articles dans cet axe.

Seul les articles des 52 UPR et UMR sont pris en compte.

4. Rédaction d'un document de synthèse de présentation de l'étude, contenant les 800 références d'articles.

Méthodologie : abrégé

Les questions :

- Est-il possible d'attribuer un axe, et un seul, à chaque article de revue ? Quelles sont les difficultés rencontrées ?

- Comment automatiser totalement ou partiellement la saisie de l'axe dans la base de données Agritrop ?
Est-il possible de classer automatiquement les articles à partir
 - du premier code Agris déjà attribué à chaque article ?
 - de l'unité de recherche des auteurs ?

- Pour chaque axe, comment les articles se répartissent-ils ?
 - selon la notoriété de la revue (facteur d'impact, comité de lecture)
 - par unité de recherche des auteurs

3. Résultats

Résultats

Est-il possible d'attribuer un axe, et un seul, à chaque article de revue ?

Quelles sont les difficultés rencontrées ?

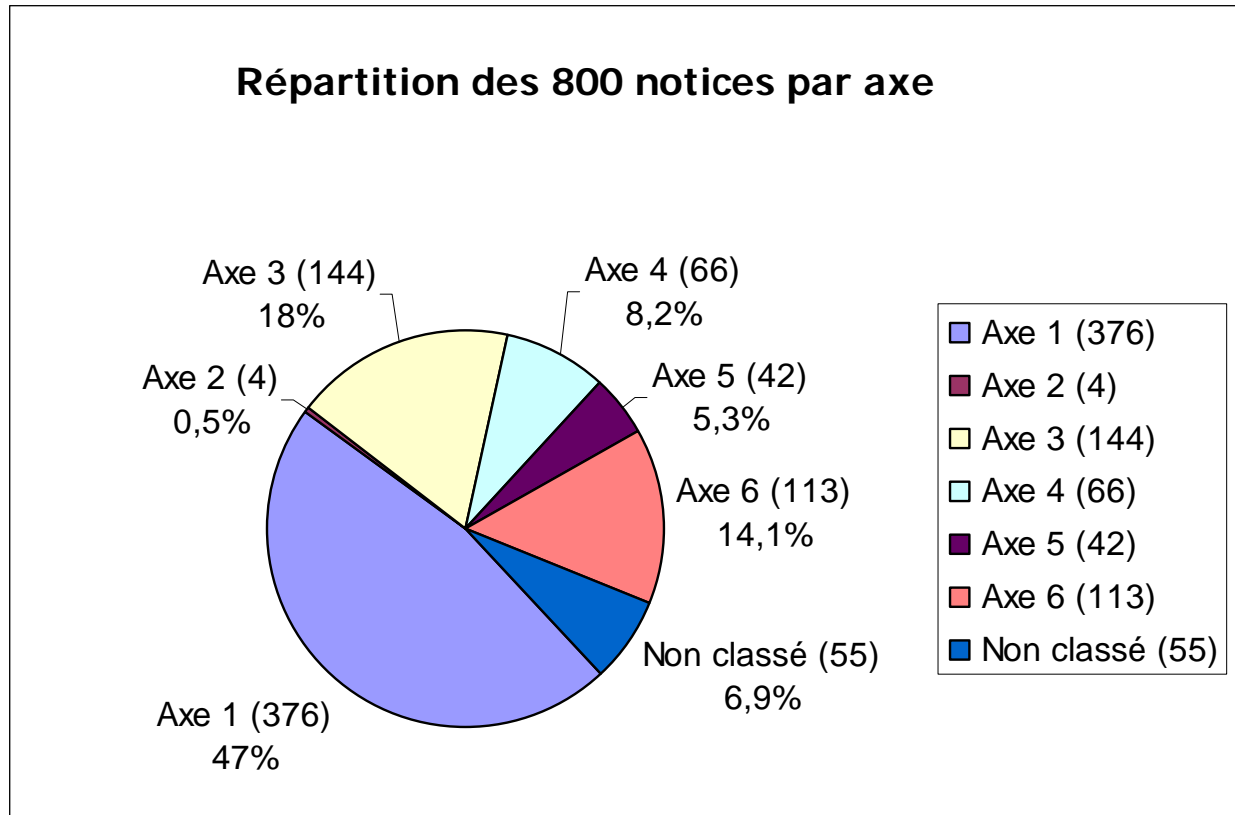
Résultats : un axe pour chaque article ?

Attribution d'un axe pour chaque article

- En général chaque article se situe dans un axe stratégique et un seul.
- La répartition du nombre d'articles entre les axes est très hétérogène.
- Seulement 6,9 % (55/800) des articles n'ont pas pu être classés dans les axes. Pour la suite de l'étude ils ont été regroupés dans une classe « non classé ».

Résultats : un axe pour chaque article ?

Attribution d'un axe pour chaque article



Faits marquants :

47 % des articles sont dans l'axe 1

0,5 % des articles sont dans l'axe 2

Résultats : un axe pour chaque article ?

Attribution d'un axe par article

- Globalement les 5 lecteurs-indexeurs classent les articles de la même façon dans les axes stratégiques.
- Quand il y a désaccord :
 - Soit l'article ne correspond pas à l'un des axes stratégiques.
 - Soit l'interprétation de la définition des axes stratégiques diverge entre les lecteurs.
- Exemple de difficultés de classement rencontrées ...

Résultats : un axe pour chaque article ?

Difficultés rencontrées

Pour certains articles, il n'y a pas d'axe stratégique correspondant.

Exemples :

- technologie des produits non alimentaires : bois, coton, caoutchouc...,
- modélisation et méthodes statistiques non appliquées,
- biochimie des lipides appliqués à des produits non alimentaire,
- étude des animaux (faunistique) ou des végétaux (forêt) sans relation avec l'agriculture ou les écosystèmes,
- changement climatique (climatologie).

Résultats : un axe pour chaque article ?

Difficultés rencontrées

Deux axes stratégiques possibles...

... sans qu'aucun ne convienne vraiment

Exemples

Entre l'axe 1 et 6 :

- étude des sols sans lien direct avec la plante, ou l'agriculture,
- gestion des ressources naturelles : eau, collection de ressources génétiques,

Entre l'axe 1 et 3 :

- étude sur les résidus de pesticides depuis le champ jusqu'à la fourchette,
- étude sur une filière.

Entre l'axe 6 et 5 :

- aménagement du territoire sans préoccupation environnementale.

Résultats

Comment automatiser totalement ou partiellement la saisie de l'axe dans la base de données Agritrop ?

Est-il possible de classer automatiquement les articles à partir du premier code Agris déjà attribué à chaque article dans Agritrop.

Résultats : classement par le code Agris (FAO) ?

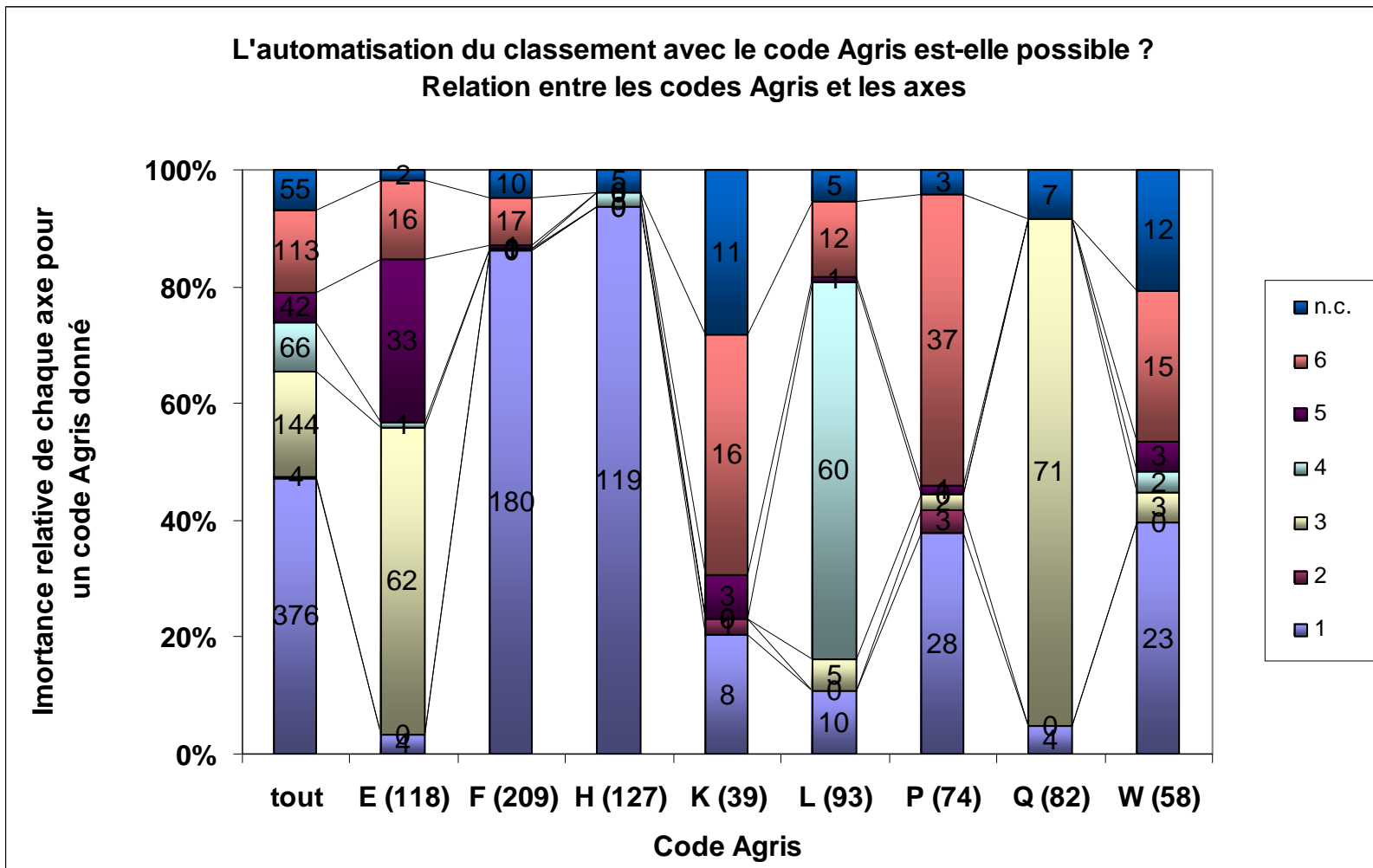
Dans Agritrop à chaque article est déjà attribué un ou plusieurs codes selon le plan de catégorisation AGRIS (FAO). Le premier code attribué est considéré comme le code principal.

Les grandes subdivisions du plan de catégorisation AGRIS (FAO) sont :

- A Agriculture en général
- B Géographie et histoire
- C Enseignement, vulgarisation et information
- D Administration et législation
- E Économie agricole, développement et sociologie rurale**
- F Sciences et production végétales**
- H Protection des végétaux**
- J Technologie après récolte
- K Foresterie**
- L Sciences animales, production et protection animales**
- M Pêche et aquaculture
- N Machinisme agricole et génie rural
- P Ressources naturelles et environnement**
- Q Traitement des produits agricoles**
- S Nutrition humaine
- T Pollution
- U Méthodologie

Compte tenu de leur faible occurrence, les articles des catégories : A B C D J N T S M ont été regroupés dans une même classe W (7 %).

Résultats : classement par le code Agris (FAO) ?



- Est-il possible de classer automatiquement les articles à partir du premier code Agris de chaque article ?
- Non pour l'ensemble du corpus.
 - Oui de façon semi-automatique pour les articles des catégories : **F** (axe 1), **H** (axe 1), **L** (axe 4), **Q** (axe 3).

Résultats

Comment automatiser totalement ou partiellement la saisie de l'axe dans la base de données Agritrop ?

Est-il possible de classer automatiquement les articles à partir de l'unité de recherche des auteurs ?

Résultats : classement automatique par UR ?

7 Unités de Recherche ont 100 % de leurs articles dans le même axe

Nature	Dpt	n° UR	Intitulé court	Axe principal	% d'articles en dehors de cet axe	nombre total d'articles	Axe 1
UPR	2	1	Couverts permanents	1	0	3	3
UPR	1	6	Peuplements riz	1	0	4	4
UPR	1	8	Agrodiversité en savane	1	0	10	10
UPR	1	29	Etiologie dépérissements	1	0	11	11
UPR	1	50	Acridologie	1	0	4	4
UPR	2	69	SCRID	1	0	10	10
UMR	1	100	DIAPC	1	0	6	6

Résultats : classement automatique par UR ?

11 Unités de Recherche ont entre 85 et 99% d'articles classés dans le même axe

							Nombre d'articles classés dans cet axe						
Nature	Dpt	n° UR	Intitulé court	Axe principal	% d'articles en dehors de cet axe	nombre total d'articles	1	2	3	4	5	6	Non classé
UMR	1	54	BGPI	1	3	35	34			1			
UPR	1	15	Contrôle des maladies	1	5	43	41					1	1
UMR	1	82	LSTM	1	6	16	15					1	
UMR	1	96	DAP	1	7	87	81		1	1			4
UPR	1	59	Modélisation intégrative	1	7	14	13		1				
UPR	1	28	Génétique Palmier	1	9	11	10		1				
UPR	1	21	Bioagresseurs	1	10	29	26				1		2
UPR	2	20	Aquaculture	1	10	10	9		1				
UPR	3	47	GREEN	6	13	24			1		2	21	
UPR	1	75	Multiplication végétative	1	14	28	24		3	1			
UMR	1	98	RPB	1	14	7	6		1				

Résultats : classement automatique par UR ?

12 Unités de Recherche ont entre 70 et 84% d'articles classés dans le même axe

Nature	Dpt	n° UR	Intitulé court	Axe principal	% d'article en dehors de cet axe	nombre total d'articles	1	2	3	4	5	6	Non classé
UMR	1	10 1	CEFE	1	17	6	5					1	
UMR	2	62	IATE%	3	18	17			14				3
UPR	2	77	Production fruitière	1	19	27	22		4				1
UMR	1	17	Trypanosomes	4	20	10	2			8			
UPR	2	34	Systèmes de pérennes	1	20	5	4					1	
UMR	2	95	QUALISUD	3	21	73	13		58				2
UMR	1	53	PVBMT	1	21	48	38			1		7	2
UPR	3	88	Politiques et Marchés	5	22	23			1		18	4	
UPR	1	39	Génétique forestière	1	22	18	14					2	2
UPR	3	70	Forêts et biodiversité	1	25	4	3					1	
UPR	3	80	Écosystèmes de plantations	1	27	26	19					5	2
UPR	2	27	Horticulture	1	27	11	8		2			1	

Résultats : classement automatique par UR ?

22 Unités de Recherche ont moins de 70% d'articles classés dans un même axe

Nature	Dpt	n° UR	Intitulé court	Axes concernés	% d'articles en dehors de l'axe majeur	nb total d'articles	1	2	3	4	5	6	Non classé
UPR	3	18	Systèmes d'élevage	1,3,4,5,6	> = 30 %	26	9		7	6	2	2	
UMR	2	84	ERRC	1,4	> = 30 %	3	2			1			
UMR	2	94	QUALITROP	1/3	> = 30 %	4	2		2				
UPR	2	102	SCA	1/4,5,3,non classé	> = 30 %	53	38		1	4	4		6
UMR	1	51	AMAP	1/6, non classé	> = 30 %	26	16					3	7
UPR	2	78	Recyclage et risque	1/6, non classé	> = 30 %	25	16					5	4
UMR	2	66	SYSTEM	1/6,3	> = 30 %	17	11		1			5	
UMR	1	55	CBGP	1/non classé, 6,4	> = 30 %	14	7			1		1	5
UPR	3	42	Biomasse énergie	2,non classé	> = 30 %	3		2					1
UPR	2	26	Systèmes bananes et ananas	3/1,6	> = 30 %	78	21		54			3	
UPR	3	22	AGIRS	4/6,1,non classé	> = 30 %	39	3			26		9	1

Résultats : classement automatique par UR ?

22 Unités de Recherche ont moins de 70 % d'articles classés dans le même axe (suite)

Nature	Dpt	n° UR	Intitulé court	Axes concernés	% d'articles en dehors de l'axe majeur	nb total d'articles	1	2	3	4	5	6	Non classé
UMR	3	99	MOISA	5/3,4,6	> = 30 %	8			2	1	4	1	
UMR	3	85	INNOVATION	5/6,1,3	> = 30 %	30	7		5		9	9	
UPR	3	68	Pastoralisme	6	un seul article	1						1	
UMR	3	93	ECOFOG	6, non classé	> = 30 %	2						1	1
UPR	3	37	Dynamique forestière	6/1, non classé	> = 30 %	27	5					19	3
UMR	3	90	G-EAU	6/1, 5 non classé	> = 30 %	20	6				1	13	
UMR	3	91	TETIS	6/1,5,4, non classé	> = 30 %	23	5			1	4	12	1
UPR	3	36	Ressources forestières	6/3,1,5, non classé	> = 30 %	37	3	1	4		3	23	3
UMR	3	56	CIREC	pas d'article		0							
UMR	2	64	GPEB	pas d'article		0							
UPR	3	40	Bois tropicaux	non classé,1,2	> = 30 %	14	3	2					9

Résultats : conclusion

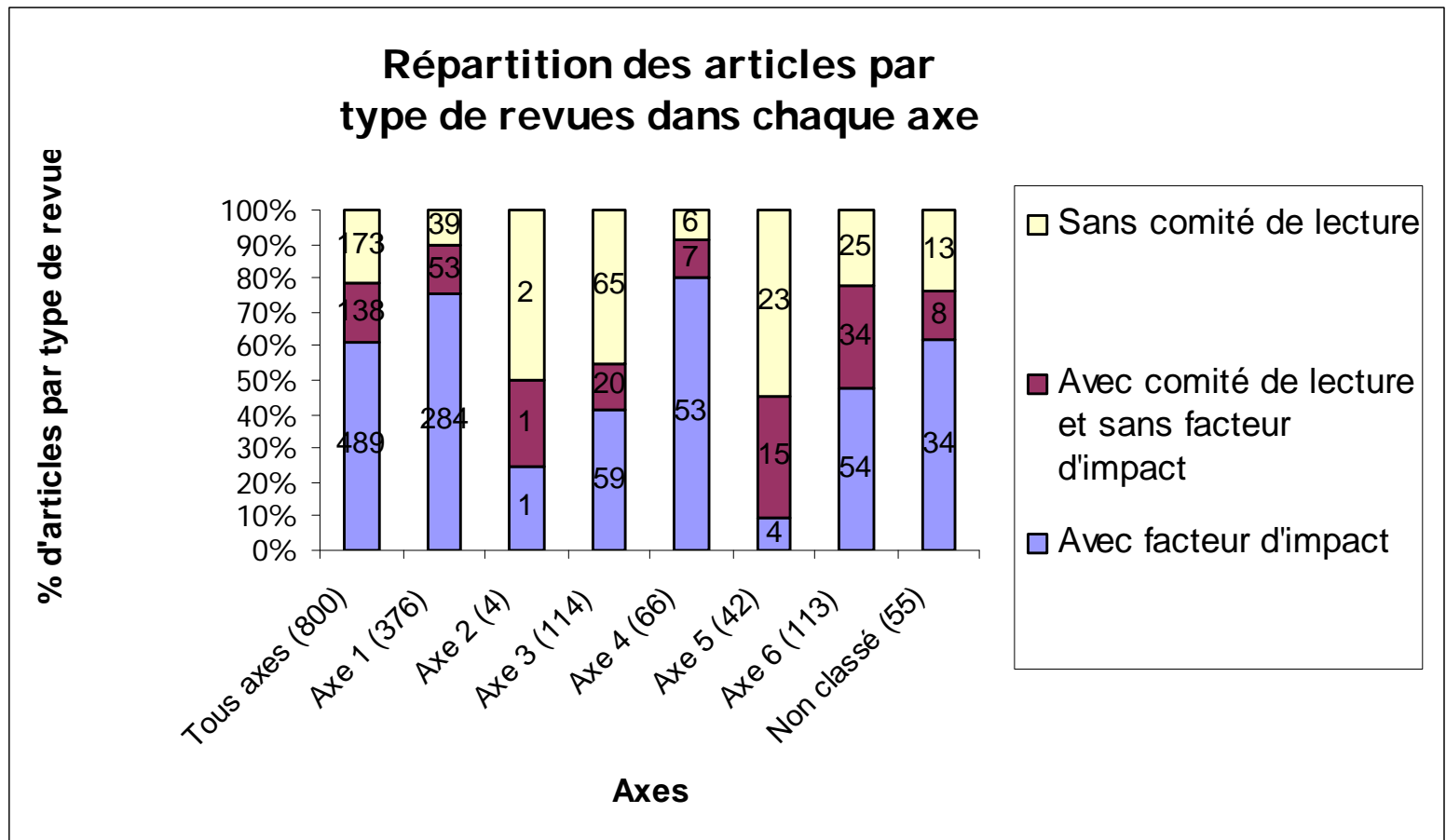
Est-il possible de classer automatiquement les articles à partir de l'Unité de Recherche des auteurs dans un axe ?

- Non pour l'ensemble des 52 UPR et UMR
- Oui pour 30 UR si l'on accepte qu'au plus 30 % de leurs articles soient en dehors de cet axe
- Oui pour seulement 15 UR si l'on accepte qu'au plus 10 % de articles soient en dehors de cet axe

3. Résultats par axe

Résultats : notoriété des revues par axe

Pour chaque axe, proportion d'articles selon la notoriété de la revue : avec facteur d'impact, avec ou sans comité de lecture



Résultats : axe 1

376 articles, soit 47% du corpus

90 % des articles dans des revues avec un comité de lecture

Les mots clés libres

AGRICULTURE DURABLE, AGROECOLOGIE, AMELIORATION DES PLANTES, CYCLE BIOGEOCHIMIQUE, SELECTION, SYSTEME DE PRODUCTION

Les unités de recherche (UMR et UPR)

45 UR (86%) ont des articles classés dans l'axe 1 :

- 25 UR pourraient être associées à l'axe 1 sur la base d'au moins 70% de leurs articles se référant à cet axe.
- 5 UR pourraient être associées à un autre axe. Elles ont cependant au moins 2 articles dans l'axe 1.
- 15 UR n'ont pas au moins 70% de leurs articles dans un même axe. Leur association à un seul axe est impossible avec ce seuil de classement de 70%.

Résultats : axe 2

4 articles, soit 0,5% du corpus

50 % des articles dans des revues avec un comité de lecture

Les mots clés libres : AGROCARBURANT, BIOENERGIE, DEVELOPPEMENT RURAL, PRODUCTION DE BIOMASSE, PRODUCTION DE L'ENERGIE, TECHNOLOGIE ENERGETIQUE

Le nombre d'articles de l'échantillon est très faible.

Les unités de recherche (UMR et UPR) :

3 UR (6%) ont des articles classés dans l'axe 2

- Aucune UR ne pourrait être associée à l'axe 2 sur la base d'au moins 70% de leurs articles se référant à cet axe.
- 3 UR n'ont pas au moins 70% de leurs articles dans un même axe. Leur association à un seul axe est impossible avec ce seuil de classement de 70%.

Résultats : axe 3

144 articles, soit 18 % du corpus

55 % des articles dans des revues avec un comité de lecture

Les mots clés libres

CONSOMMATION, FILIERE AGRO-ALIMENTAIRE, LEGISLATION ALIMENTAIRE, PRODUIT VIVRIER, QUALITE, SURETE ALIMENTAIRE, TECHNOLOGIE ALIMENTAIRE

Les unités de recherche (UMR et UPR)

20 UR (38%) ont des articles classés dans cet axe :

- 2 UR pourraient être associées à l'axe 3 sur la base d'au moins 70% de leurs articles se référant à cet axe.
- 11 UR pourraient être associées à d'autres axes. Elles ont cependant au moins 1 articles dans l'axe 3.
- 7 UR n'ont pas au moins 70% de leurs articles dans un même axe. Leur association à un seul axe est impossible avec ce seuil de classement de 70%.

Résultats : axe 4

66 articles, soit 8,2% du corpus

91 % des articles dans des revues avec un comité de lecture

Les mots clés libres

CONTRÔLE DE MALADIES, EPIDEMIOLOGIE, MALADIE A VECTEUR, MALADIE ANIMALE INFECTIEUSE, RELATION HOTE-PATHOGENE-VECTEUR, RISQUE SANITAIRE, ZOOOSE, (MALADIE ANIMALE)

Les unités de recherche (UMR et UPR)

13 UR (25%) ont des articles classés dans l'axe 4 :

- 1 UR pourraient être associées à l'axe 4 sur la base d'au moins 70% de leurs articles se référant à cet axe.
- 11 UR pourraient être associées à d'autres axes. Elles ont cependant au moins 1 articles dans l'axe 4.
- 1 UR n'a pas au moins 70% de ces articles dans un même axe. Son association à un seul axe est impossible avec ce seuil de classement de 70%.

Résultats : axe 5

42 articles, soit 5,3 % du corpus

45 % des articles dans des revues avec un comité de lecture

Les mots clés libres

ACTION PUBLIQUE, AGRICULTURE FAMILIALE, LIBERALISATION DES ECHANGES, OBJECTIF DU MILLENAIRE, PAUVRETE ET INEGALITE, PLANIFICATION, POLITIQUE AGRICOLE, REGULATION DE LA PRODUCTION

Les unités de recherche (UMR et UPR)

10 UR (19%) ont des articles classés dans l'axe 5 :

- 1 UR pourrait être associée dans l'axe 5 sur la base d'au moins 70% de leurs articles se référant à cet axe.
- 2 UR pourraient être associées à un autre axe. Elles ont cependant au moins 2 articles dans l'axe 5.
- 7 UR n'ont pas au moins 70 % de leurs articles dans un même axe. Leur association à un seul axe est impossible avec ce seuil de classement de 70%.

Résultats : axe 6

113 articles, soit 14,1% du corpus

78 % des articles dans des revues avec un comité de lecture

Les mots clés libres

AMENAGEMENT DU TERRITOIRE, BIODIVERSITE, ECOSYSTEME, GESTION DES RESSOURCES NATURELLES, GOUVERNANCE, IMPACT ECOLOGIQUE, PAYSAGE, RELATION SOCIETE HUMAINE ET NATURE

Les unités de recherche (UMR et UPR)

26 UR (50%) ont des articles classés dans l'axe 6 :

- 1 UR pourraient être associées à cet axe sur la base d'au moins 70% de leurs articles se référant à l'axe 6.
- 10 UR pourraient être associées à un autre axe. Elles ont cependant au moins 2 articles dans l'axe 6.
- 15 UR n'ont pas au moins 70 % de leurs articles dans un même axe. Leur association à un seul axe est impossible avec ce seuil de classement de 70%.

Résultats : non classé

55 articles, soit 6,9 % du corpus

76 % des articles dans des revues avec un comité de lecture

Quelques thèmes des articles non classés

- technologie des produits non alimentaires (sauf biomasse énergie) : bois, coton, caoutchouc...,
- étude des animaux (faunistique) ou des végétaux (forêt) sans relation avec l'agriculture ou les écosystèmes,
- utilisation des végétaux pour la santé humaine,
- modélisation non appliquée sans lien avec le végétal, l'animal ou l'agriculture,
- climatologie.

Les unités de recherche (UMR et UPR)

20 UR (38%) ont des articles non classés :

- 9 UR pourraient être associées à des axes. Elles ont cependant au moins 1 article « non classé ».
- 11 UR n'ont pas au moins 70 % de leurs articles dans un même axe. Leur classement dans un seul axe est impossible avec ce seuil de classement de 70%.

Conclusions

Résultats : conclusion

Les réponses aux questions pour les articles de revue de ce corpus

- Est-il possible d'attribuer un axe et un seul à chaque article de périodique ? Quelles sont les difficultés rencontrées ?
 - Oui avec un taux de 6,9 % d'articles non classés.
- Comment automatiser totalement ou partiellement la saisie des axes dans la base de données Agritrop ?

Est-il possible de classer automatiquement les articles à partir :

- du premier code Agris attribué à chaque article ?
 - Non pour l'ensemble
 - Oui de façon semi-automatique pour les articles classés des catégories : F, H, L, Q.
- de l'unité de recherche des auteurs ?
 - Non pour l'ensemble des 52 UPR et UMR
 - Oui pour 30 UR si l'on accepte que 30 % de leurs articles soient en dehors de cet axe.
 - Oui pour seulement 15 UR si l'on accepte qu'au plus 10 % de leurs articles soient en dehors de cet axe .

Conclusion

- Cette étude expérimentale ne concerne que les articles 2007-2008 enregistrés avant le 31/05/2008. Les résultats ne sont qu'indicatifs.
- Un document papier contenant les 800 références bibliographiques, classées par axe, est joint en annexe.
- Suite à cette étude, dans Agritrop , il est possible d'attribuer un axe stratégique à chaque article de revue, à partir de l'année de publication 2007.
- La mention de l'axe est une nouvelle information mémorisée et réutilisable par tous, qui pourra être utilisée pour construire des indicateurs.
- Pour le contrat d'objectifs 2008-2011 il est envisagé d'ajouter cette information pour les articles de revue 2005-2006 (avec une échéance en avril 2009). La même méthodologie d'attribution des axes sera utilisée. La saisie dans Agritrop sera semi-automatisée à partir du code Agris.