



**HAL**  
open science

# Les effets d'une programmation thématique des activités de la recherche publique. Résultats d'une étude socio-sémantique des projets soumis aux programmes ANR ADD, SYSTERRA et AGROBIOSPHERE

Marc Barbier, Philippe Breucker

## ► To cite this version:

Marc Barbier, Philippe Breucker. Les effets d'une programmation thématique des activités de la recherche publique. Résultats d'une étude socio-sémantique des projets soumis aux programmes ANR ADD, SYSTERRA et AGROBIOSPHERE. [Rapport Technique] Agence Nationale de la Recherche. 2015, 52 p. hal-02508168

**HAL Id: hal-02508168**

**<https://hal.science/hal-02508168>**

Submitted on 13 Mar 2020

**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.

RAPPORT ANR  
UNE ÉTUDE D'IMPACT DES PROGRAMMES  
ADD-SYSTERRA-AGROBIOSPHERE, MENÉE EN  
COLLABORATION AVEC L'UNITÉ DE RECHERCHE  
INRA SENS (UMR 1326).

LES EFFETS D'UNE PROGRAMMATION THÉMATIQUE  
DES ACTIVITÉS DE LA RECHERCHE PUBLIQUE.

RÉSULTATS D'UNE ÉTUDE SOCIO-SÉMANTIQUE DES PROJETS SOUMIS  
AUX PROGRAMMES ANR ADD, SYSTERRA ET AGROBIOSPHERE.

Publié en juillet 2015

Auteurs : Marc Barbier\* et Philippe Breucker\*

En collaboration avec Emmanuelle Lemaire\*\* et Florence Jacquet\*\*

\*INRA, UMR 1326 LISIS, F 77454 Marne-La-Vallée, France

\*\*ANR, Paris, France

Ce travail a été réalisé en lien avec le cahier de l'ANR n°8  
Émergence de l'agroécologie et perspectives pour le futur : Les programmes  
ADD-SYSTERRA-AGROBIOSPHERE - publié en septembre 2015

À l'issue de 8 années de programmation, l'ANR a initié une réflexion sur le bilan et les impacts des programmes ADD/SYSTERRA/AGROBIOSPHERE et sollicité la conduite d'une étude sur les projets soumis et lauréats de ces programmes.

Cette communication propose les résultats d'une entreprise de caractérisation de ces impacts fondée sur l'analyse de données textuelles de corpus de projets répondant aux appels de trois programmes successifs. Suivant une méthodologie d'étude de domaine de recherche déjà éprouvée, le travail a consisté à caractériser l'évolution conjointe des thématiques de recherche et la structuration de communautés de recherche, suivant l'idée de la structuration d'une offre de promesses scientifiques portées par des collaborations de laboratoires.

Les analyses ont été possibles grâce à la formation d'une base de données des projets soumis qui a fait l'objet d'un important travail ingénierique (application projbase). Ces analyses ont été réalisées avec la plateforme CorText grâce à l'application CorText-Manager qui permet des calculs de graphes de mots-associés et la visualisation de clusterisation de ces graphes.

Ce travail exploratoire correspond à un intérêt grandissant de communautés en sciences sociales pour les dynamiques de recherche et les politiques qui les visent à un moment où les grands enjeux du développement durable font l'objet d'un traitement accru dans le cadre d'une gouvernance naissante du système terre.

Contact : [marc.barbier@grignon.inra.fr](mailto:marc.barbier@grignon.inra.fr)

## SOMMAIRE

<b>Introduction .....</b>	<b>3</b>
<b>1. Présentation de l'étude.....</b>	<b>4</b>
1.1. Etat de l'art.....	4
1.2. Problématique scientifique : le projet comme micro organisation-frontière.....	6
1.3. Méthodologie .....	7
<i>Considérations générales</i> .....	7
<i>Création de la base de données des projets</i> .....	7
<i>Plan de l'analyse</i> .....	8
<b>2. Résultats.....</b>	<b>11</b>
2.1. L'offre de programme : les appels comme attracteurs lexicaux.....	11
2.2. Analyse structurale du paysage sémantique des projets soumis.....	15
2.2.1. <i>Démographie des multi-termes principaux</i> .....	15
2.2.2 <i>Le paysage sémantique global des multi-termes</i> .....	15
2.3. Analyse de la dynamique temporelle des multi-termes .....	19
2.4. Le paysage des collaborations entre unités de recherche .....	24
2.4.1. <i>Le sens des assemblages</i> .....	24
2.4.2. <i>Un mouvement d'extension des collaborations</i> .....	24
2.4.3. <i>La dualité de l'organisation par projets</i> .....	26
2.4.4. <i>La composition socio-sémantique des collaborations</i> .....	28
<b>Conclusion.....</b>	<b>37</b>
Références bibliographiques .....	40
<b>Annexes</b>	
Annexe 1. Présentation de la plateforme CorTexT .....	44
Annexe 2 : Clusterisation et tableaux des N-Gram extraits des textes des Appels à Projets.....	45
Annexe 3 : clusterisation des termes des résumés pour les 3 programmes.....	46

## Introduction

A l'issue de huit années de programmation, l'ANR a initié une réflexion sur le bilan et les impacts des programmes ADD/SYSTERRA/AGROBIOSPHERE et sollicité l'unité Inra-SenS pour conduire une étude sur les projets soumis et lauréats de ces programmes. Cette étude est d'abord une entreprise de caractérisation fondée sur l'analyse de données textuelles de corpus de projets répondant aux appels de trois programmes successifs. Elle a pour objectif d'explorer l'impact de ces programmes en termes de l'évolution des thématiques de recherche et de structuration de communautés de recherche. Elle correspond à un intérêt grandissant de communautés en sciences sociales pour les dynamiques de recherche et les politiques qui les visent à un moment où les grands enjeux du développement durable souligné dès la conférence de Rio en 1992 font l'objet d'un traitement accru dans le cadre d'une gouvernance naissante du système terre. Pour la recherche agronomique, elle-même en pleine recomposition de ces domaines et de ses fronts de recherche, la période de programmation dont il est question ici se trouve à un moment charnière, tant du point de vue de la constitution de capacités de recherche sur et pour la contribution de l'agriculture au développement durable que du point de vue de la transformation des modalités de soutiens publics à la recherche par la création de l'Agence Nationale de la Recherche.

La réalisation de cette étude a été accompagnée par un groupe de suivi composé autour du comité scientifique du Colloque ANR 2014, et elle a été menée conjointement par l'Inra-Sens et l'ANR. La présente communication restitue les résultats de cette étude qui ont nourri une présentation réalisée au colloque de l'ANR en Décembre 2014 et un des chapitres des Cahiers de l'ANR dont l'édition a suivi ce colloque (ANR, 2015). C'est donc une étude qui a bénéficié d'interactions dans sa conduite et dans la définition des questions posées. Ce n'est pourtant pas le point de vue de l'Agence qui s'exprime dans cette contribution, mais bien un point de vue de chercheurs sur les projets à travers lesquels les communautés de recherche ont proposé d'engager leurs compétences et leurs capacités.

Cette étude propose tout d'abord une mise à plat de son propre positionnement scientifique, tant du point de vue de l'analyse des mots-associés que de l'analyse sociologique des projets de recherche, suivi de la présentation de la problématique de recherche et de la méthodologie en insistant sur les aspects techniques de la mise en œuvre des moyens nécessaires à la constitution des corpus. Elle restitue ensuite les principaux résultats des analyses conduites : tout d'abord concernant l'analyse de l'offre programmatique de façon à prendre la mesure de ce qui a été proposé aux équipes de recherche ; ensuite elle restitue un premier niveau de lecture des projets soumis aux appels pour caractériser le paysage sémantique d'une offre de recherche globale; puis elle propose une lecture de la dynamique temporelle de cette offre pour en caractériser l'évolution; enfin elle restitue une analyse socio-sémantique des collaborations entre laboratoires de recherche pour établir la dynamique sociale de cette évolution. En conclusion elle établit un certain nombre de faits saillants pour éclairer les dimensions de l'impact de ces programmes en termes d'évolution des thématiques et de structuration de communautés de recherche.

## 1. Présentation de l'étude

### 1.1. Etat de l'art

L'analyse des données fournies grâce à cette étude repose sur un travail de type scientométrique, dans la lignée des travaux ouverts par des approches dite d'analyse des mots associés (ou *co-word analysis*), c'est à dire un type d'analyse reposant sur la valorisation des associations contenues dans la structure des champs de bases de données textuelles et dans les contenus de ces champs quand ils font l'objet d'une extraction terminologique. La *Co-word analysis* repose fondamentalement sur l'analyse de graphes d'associations, analyse qui prolonge les approches pionnières de co-citation dans les articles scientifique (Small, 1973 ; Price, 1976). En déplaçant l'analyse vers les contenus d'autres champs que ceux des auteurs ou des références citées, elle a été largement inspirée par la théorie dite de l'acteur-réseau (Callon et al. 1983) et l'implémentation d'algorithme de cartographie de clusters de termes, d'entités nommées ou de groupes nominaux extraits automatiquement de champs ou de textes pleins (He, 1999). Elle est certes apparue dans un contexte d'accompagnement de la formulation et de l'évaluation de politiques scientifiques (Callon et al. 1986; Law et al. 1988 ; Rip 1988), mais elle a beaucoup contribué à doter les chercheurs en sciences sociales de capacités de représentations et de lectures des associations existantes, y compris pour fonder l'étude de leur propre domaine (voir par exemple de Saint Léger et van Meter, 2005).

L'analyse sociologique des réseaux de la science, fondée à l'origine sur l'étude des bases de notices bibliographiques, a connu un développement important avec le déploiement de l'évaluation à partir d'outils et de métriques de mesure d'impact (par exemple Van Raan, 2000 ; Ding et al., 2001) et de cartographies de la science (Börner et al., 2003). Pour autant que les métriques de l'évaluation soient l'objet de critiques fortes, produites tant par la communauté de recherche en sciences sociales qui les étudient (Gingras, 2008; Pontille and Torny, 2010), que par les chercheurs eux-mêmes qui reçoivent et construisent cette évaluation par leurs productions, tout un ensemble de travaux se sont particulièrement développés depuis une vingtaine d'années autour de l'analyse des réseaux de la recherche tels que différentes bases de données permettent de les caractériser (Bar-Ilan, 2008). Un projet d'investigation plus large s'est ainsi constitué au croisement de l'étude de différentes bases de données qui tracent l'activité scientifique et technique: bases de notices bibliographiques, bases de brevets, bases de projets et autres données liées aux activités de communication des chercheurs ou de leurs institutions sur le web ou dans les réseaux sociaux. Avec des perspectives plus herméneutiques, utilisant elles-aussi une lecture des associations (Chateauraynaud et al., 2003) dans une pratique de création/interprétation de composition des corpus, l'analyse des réseaux de termes a ainsi apporté un renouvellement à une génération d'outils issus la statistique textuelle (Jenny, 1997).

La *co-word analysis* a reçu des critiques, notamment sur la signification de l'extraction de mots hors de leur contexte d'énonciation (Leydesdorff and Hellstein 2006). Mais il s'agit de noter que, depuis une dizaine d'années, se sont développés énormément de projets de recherche et de technologies qui utilisent largement la théorie des graphes et l'analyse de réseau (projet Ucinet, Pajek et Gephi, Citespace notamment), avec justement l'hypothèse que les relations sont des contextes d'énonciation produits en même temps que les traces de l'activité. Cette mobilisation de la théorie des graphes s'accompagne d'un travail de plus en plus important pour analyser des dynamiques d'association de termes ou de multi-termes (N-gram) dans le temps (par exemple Cointet et Chavalarias, 2013) et pour déployer des efforts importants en visualisation des graphes (Börner et al, 2003). Ce déploiement conduit à des formulations de

méthodologies ou d'approches amples de l'ensemble de réseaux constitutifs de systèmes de connaissances (Heimeriks et al., 2003). Il forge également des travaux de sociologie des sciences reposant sur la caractérisation des relations de collaborations dans la science (Powell et al. 2005 ; Lazega , et al., 2004; Cambrosio et al. 2004; Cambrosio et al. 2006; Bourret et al. 2006) et sur celle de l'émergence de domaine de recherche multi ou transdisciplinaire (Lucio-Arias and Leydesdorff 2007), de recherche dite « translationnelle » (Jones et al. 2011) ou de fronts de recherche liés à des technologies émergentes (Mogoutov and Kahane 2007 ; Ding et al., 2001) ainsi qu'à des enjeux de durabilité plus proche du domaine que nous étudions ici (Barbier et al., 2012 ; Tancoigne and Barbier et al., 2014).

Tracer, représenter et caractériser la dynamique des savoirs dans différentes sources de données devient un enjeu important, tant sur le plan scientifique que technologique. Cet enjeu trouve dans la prolifération des données ouvertes tout un ensemble de possibilités qui sortent les chercheurs, comme leurs institutions, d'un rapport contrit ou seulement critique aux données bibliométriques. En effet, il existe de nombreux jeux de données dans les organisations de recherche ou les agences qui financent ou évaluent les chercheurs et leurs organisations, sans compter les espaces d'échange collaboratif académique sous forme de réseaux sociaux professionnels. Le potentiel de données est donc important, le croisement de ces données posent de nombreuses difficultés techniques et constituent des verrous pour la recherche en informatique et en analyse de données.

L'accélération de la production des données directement liées à différents pans de l'activité scientifique est importante. Elle suit aussi la généralisation de politiques de la recherche et de l'innovation fondées sur des programmes pilotés au niveau national ou européen (Larédo et Mustar, 2001) à travers un management par projets bien connus des chercheurs (Laudel, 2006). Si la France a connu une politique de la recherche et de l'innovation par programme au sein des établissements publics avant la création de l'ANR, c'est avec la naissance de cette agence que des capacités publiques de financement thématiques ont été largement instituées, au point que le financement par appels d'offre en devienne structurant (Thèves et al. 2007) : il devient alors difficile de distinguer dans les promesses scientifiques que portent les projets ce qui relève de la science à faire ou des affaires de science. L'analyse de cette couche d'information constituée par des bases de données dont les projets sont les entités est alors un enjeu de recherche pour les sciences sociales.

Un certain nombre de travaux récents à la frontière de la sociologie des organisations et de la sociologie des sciences indiquent qu'un mouvement de transformations professionnelles est à l'œuvre dans les organisations de recherche en quête de performances (Paradeise & Thoenig, 2013), notamment à travers les effets de l'organisation de l'activité par projets. La recherche sur ces effets est en plein essor: effets sur l'organisation du travail des chercheurs au sein d'une matrice qui croise la vie des équipes, la vie de l'établissement et des univers de réseaux scientifiques (Louvel, 2007; Hubert et Louvel, 2012 ; Tari, 2015) ; effets sur le temps du travail scientifique avec la temporalité particulière qu'instigie le temps du projet (Schultz, 2013) ; effets croisés de la composition conjointe des équipes, des domaines de recherche et de la recomposition des relations avec les partenaires industriels (Barrier, 2010, 2011) ; effet sur le sens donné par les chercheurs à ces recompositions à la fois politiques et managériales (Hubert et al., 2012) et les conséquences en termes de redéfinition de la professionnalité (Jouvenet, 2011). A ce déploiement de recherches qui prolongent les questionnements nés avec les *science and technology studies* il y a plus de 20 ans, s'ajoutent donc celles, plus proches des *Sciences Policy Studies* qui portent sur les bases de données qui tracent la gestion des activités de recherche à différents niveaux d'organisation (par exemple Cordis, JOREP, EUPRO au niveau européen). Les données sur les projets peuvent alors, à leur tour, faire l'objet d'études bibliométrique dans une quête d'indicateurs pour la politique et l'évaluation de la recherche (Laredo, 2010 ; Lepori et al, 2007) ou servir à des entreprises d'analyse et de modélisation des collaborations scientifiques telles que décrites dans ces projets, avec la question de savoir si les projets font, ou bien résultent, des collaborations (Tari et al., 2010 ; Hoekman, et al., 2013) et ce qu'il peut résulter

de cette organisation des collaborations dans la publication scientifique en terme d'impact (Katz et Martin, 1997; Landry et al., 1999 ; Lee and Bozeman, 2005). L'idée que la formation de projets en réponse à des appels d'offre puissent être uniquement le résultat d'une incitation financière en vue de publier est ainsi contrastée par l'étude empirique de la pluralité des motivations (Beaver, 2001) et peut conduire à distinguer ce qu'un projet représente pour les financeurs d'un côté et pour les chercheurs de l'autre (Bozeman and Rogers, 2002). Cet aspect renforce l'idée d'une dualité de l'agencement et de la structure et des transformations dont cette dualité est le moteur (Sewell, 1992) ; elle doit alors devenir un objet d'investigation pour les recherches sur l'organisation et la gouvernance des activités scientifiques et techniques. Notre étude est ainsi marquée par l'hypothèse de l'existence d'une telle dualité, et des effets qu'elle peut engendrer de façon contingente ou déterministe.

## 1.2. Problématique scientifique : le projet comme micro organisation-frontière

Plutôt que de vouloir faire la critique de cette dualité et des découplages entre l'agencement des collaborations scientifiques et la structure des politiques et des financements de la recherche, il semble d'abord pertinent de la considérer comme un objet d'étude, et eux-mêmes en travail dans les organisations. Partant, on peut faire l'hypothèse que le projet de recherche est une micro organisation-frontière, manifestation contingente d'une dynamique des acteurs de la science en action au sein de cette dualité organisationnelle, micro-organisation sporadique du point de vue de sa gestion mais qui repose sur des relations existantes avant et après.

A la constitution de cette micro organisation-frontière correspond une production documentaire importante, réalisée par des collectifs de chercheurs à la recherche de financements sous une forme assez standardisée, avec des formats qui fondent un genre particulier de littéracie scientifique (Miller, 1983), c'est à dire une façon d'écrire et de proposer les fins, les moyens et l'organisation de l'activité scientifique, sans que le document ne soit une production académique ou de « vulgarisation ». C'est une forme particulière d'écriture dédiée aux décideurs comme aux pairs, que l'on peut penser assez spécifique au format du projet de recherche et qui mérite d'être étudiée (Cetro et al., 2014). Car un projet de recherche est une promesse hybride fondée sur plusieurs ressources argumentaires: sur la recherche-à-faire à la suite de la recherche-déjà-faite ; sur la réalité et la félicité d'un collectif de personnes à la suite souvent d'expériences de travail en commun précédentes ; sur l'organisation spécifique du travail en tâches supposant une rationalisation possible ou effective du travail de recherche ; sur la disponibilité des ressources et des capacités déjà-là qui viennent justifier l'intérêt de financer des équipes. Ainsi, un projet de recherche sur appel d'offre n'est pas seulement un document pour capturer des fonds puis assumer, à travers un rapport scientifique et opérationnel, la reddition de compte attendue du financeur : c'est aussi une promesse ou une fiction rationnelle ancrée dans la réalité de la production scientifique déjà réalisée et qui met en action des collectifs sur la base d'un schéma d'organisation.

Pris sous cet angle, la perspective ouverte par le traitement des bases de données de projets est selon nous plus importante qu'il n'y paraît à premier vue dans le registre d'une critique du managérialisme. On défend ici l'importance de pouvoir accéder à une étude de cette forme d'écriture scientifique sans la replier sur des considérations purement linguistiques, on cherchera à la considérer plutôt comme un travail de *self-accountability* de l'organisation de la recherche par les promesses. Il s'agit alors de coupler l'étude de bases de données de projets avec l'étude sociale de l'organisation par projet dans un milieu scientifique dont le programme qui fédère et finance ces projets indexe le périmètre. C'est dans cette perspective que nous avons ouvert cette étude sur trois programmes successifs de l'ANR qui ont marqué un tournant dans le monde de la recherche agronomique en France.

### 1.3. Méthodologie

Notre travail s'inscrit dans une visée compréhensive des dynamiques de la recherche publique, et non directement évaluative ; il repose sur la mobilisation de techniques informatiques et d'outils de visualisation et d'analyse de données textuelles développés par la plateforme CorTexT (annexe 1). Il s'agit d'un instrument se présentant sous la forme d'une interface accessible en ligne (<http://manager.cortext.net/>) qui contribue au renouvellement des analyses de réseaux de traces numériques, notamment celles, nombreuses, qui accompagnent les dynamiques de sciences et d'innovation en société (Barbier et Cointet, 2012).

#### Considérations générales

S'appuyant sur le déploiement d'algorithmes nouveaux (Chavalarias and Cointet 2008; Cointet, 2009 ; Roth and Cointet 2010) incluant notamment des analyses de réseaux hétérogènes (Cambrosio et al., 2010) l'application CorTexT.Manager apporte des capacités nouvelles dans le cadre du projet de la Plateforme CorTexT. De plus, des capacités de traitement et de fouilles de données sont présentes sur cette plateforme pour rendre possible le traitement de tout type de corpus textuel peu ou très calibré. Jointe à la discussion des méthodes de la *co-word analysis*, et plus largement de la mobilisation de l'analyse de réseaux, un espace de débat et de conception s'ouvre avec l'utilisation des possibilités de cartographie des connaissances dans des contextes de décision ou de formulation stratégique dans les mondes professionnels de la recherche. Plusieurs plateformes proposent ainsi des capacités équivalentes d'analyse (plateforme de l'ISC-PIF, MediaLab à SciencePo Paris, à Lyon, notamment). Dans ces contextes, les analyses de réseaux trouvent bien d'autres fonctions que celle de l'évaluation.

Mobiliser l'application CorTexT.Manager dans le contexte d'usage qu'offre cette étude avec l'ANR, repose sur une expérimentation collective avec des chercheurs et des acteurs des politiques de la science bien distincte de celle de l'évaluation par facteur d'impact. Elle correspond à un déplacement des instruments issus de la recherche vers des situations d'énonciation des politiques des sciences (Noyons, 2001). C'est un travail délicat car il allie le *design* de dispositif de production de données et d'analyse, à celui de leur mise en usages dans ce type de contexte. Les membres de la plateforme CorTexT ont déjà eu quelques expériences de ce type, objet de publications (Barbier et al. 2012, 2012; Raimbault et al., 2014; Tancoigne, Barbier et al., 2014). Elles ont permis de mettre au point des méthodes d'intervention et d'usage des outils de la dite plateforme, non seulement au niveau de la conduite des analyses, mais surtout en amont sur la construction des jeux de données, point sensible pour la qualité des analyses attendues, mais aussi pour les réflexions d'une organisation centrée sur la gestion des incitations faites à la recherche.

#### Création de la base de données des projets

Une base de données de l'ensemble des projets soumis aux trois programmes a été constituée à partir des informations disponibles au niveau du système de gestion des trois appels à projets successifs. Le fait d'accéder aux projets soumis est d'une importance majeure, puisque la population de projets que nous récoltons est porteuse d'une information bien différente de celle que forment les bases de données de projets lauréats. C'est un accès direct à la production documentaire des communautés de recherche, le fait que les projets aient été retenus ou pas devient dès lors une variable supplémentaire.

Cela a nécessité un important et long travail de recollement et de structuration d'une base de données dont les entités clés sont les projets. Une application nommée *ProjBase* développée dans le cadre du projet ANR PanBioptique (Barbier, et al., 2013) a permis de développer une approche rigoureuse pour constituer une base de données relationnelle dans laquelle les projets sont décrits par de nombreux champs issus du travail de recollement (Figure 1). Cette application a été prise en main dans le cadre du projet avec ANR en tant

que prototype pour appréhender une gestion robuste des données et pouvoir mettre en place une réflexion sur l'impact des programmes. Des enseignements en sont attendus.

Les analyses réalisées dans le cadre de cette étude ont été produites à partir des jeux de données suivantes ; c'est un choix limitatif par rapport à l'ensemble des descripteurs possibles, mais la couverture est exhaustive grâce à un important travail de nettoyage, de recollement et dé-doublonnage.

- Les projets analysés soumis aux 7 appels d'offres avec leurs descripteurs;
- Les titres et résumés de ces projets sont l'objet d'une extraction terminologique ;
- L'organisme du coordinateur et des participants pour l'information sur les « labos » porteurs.

Au sein de cette application *ProjBase*, via une capacité d'exportation paramétrables, des fichiers peuvent être produits pour constituer des jeux de données pour en conduire leur analyse avec l'interface en ligne CorText-Manager, sachant que d'autres outils pourraient s'en saisir puisque le format de sortie est de type tabulaire.

### Plan de l'analyse

Une fois ce travail réalisé, l'objectif était alors de caractériser la structure et la dynamique d'un espace scientifique à la fois porté par les communautés de recherche et induit par l'incitation répétée à trois reprises par l'ANR pour orienter ces communautés sur les thématiques des programmes. Nos premières analyses de cette population de projets permettent d'instruire les deux objets principaux d'enquêtes qui font l'objet du présent document:

- l'évolution des thématiques traitées par les projets soumis et financés à partir d'une détection de motifs lexicaux et à une cartographie de leur distribution temporelle ;
- les réseaux sociaux et institutionnels des partenariats de projet à partir des affiliations des participants.

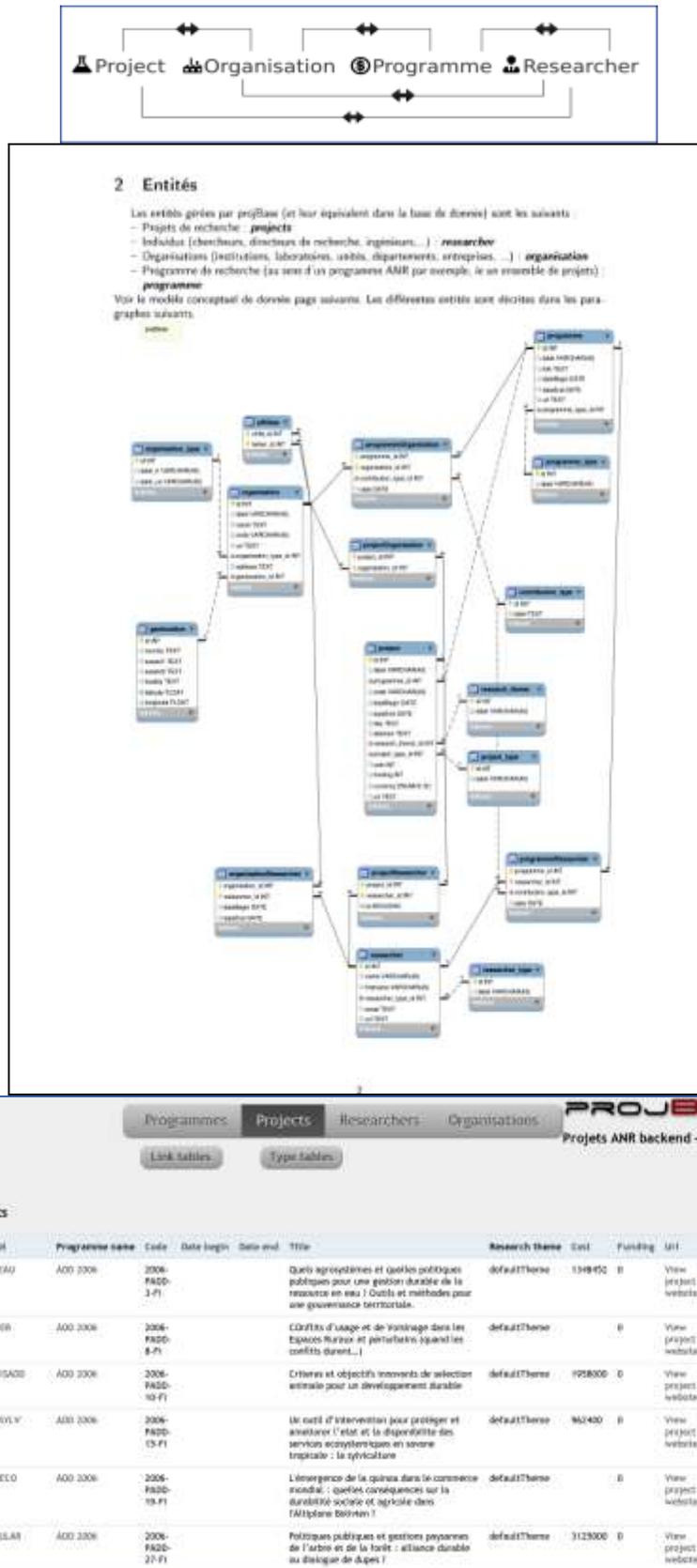
Les champs descripteurs de ces projets ont pu être facilement indexés et, pour les champs textuels riches, une extraction terminologique a été conduite (seul le champ résumé est étudié à ce stade, mais d'autres champs sont en cours de traitement: texte sur la problématique de recherche, texte sur le positionnement par rapport à la thématique). Cette extraction terminologique vise le repérage de groupes nominaux fait de plusieurs termes (N-gram) porteurs par leur association fréquente d'une information. A partir de toute sélection de ces n-gram (au seuil  $n > 2$  termes), la constitution de graphes des relations suivie d'un calcul de détection de communautés permet de visualiser la distribution de ces « multi-termes » et leur agencement en clusters, caractérisant des zones denses de discours tenus dans les résumés des projets. Les analyses ont été faites sur les 150 premiers « multi-termes » (en fréquence d'apparition) du fait de la taille faible de la base et de la distribution des fréquences de termes.

L'évolution temporelle est introduite à partir de période recouvrant les 3 programmes (ADD, SYSTERRA et AGROBIOSPHERE) pour identifier des tendances (émergence, continuité ou disparition des clusters) et plusieurs types de visualisation de données proposées par l'application CorText-Manager ont été utilisées pour permettre un travail interprétatif: description des fréquences d'occurrence des termes, réseaux des multi-termes sur l'ensemble des programmes (paysage général), projection des différents programmes sur ces réseaux, réseaux de laboratoires liés par leurs contributions aux projets.

La lecture de ces graphes présentés ci-après dans le corps du texte, mérite quelques explications: la taille des nœuds (triangle) est proportionnelle au nombre de liens que ce nœud possède. La taille des liens est proportionnelle au nombre de liens entre les 2 nœuds qu'il relie. La taille du cluster est proportionnelle au nombre de projets qui contribuent au cluster. Les cartes ont été construites à partir d'une mesure indirecte

de similarité dite « Distributionnal » pour les graphes homogènes (à savoir que les noeuds ont des valeurs prises dans un même champ), et à partir d'une mesure directe de similarité dite du « Chi2 » pour les graphes hétérogènes (les nœuds ont des valeurs prises dans deux champs différents). Enfin, un algorithme de détection de communauté permet de calculer ces clusters et leur visualisation utilise l'algorithme, le plus standard et reconnu comme le plus robuste : « Louvain ».

Figure 1. Schéma de principe de la base de données établie avec l'application ProjBase



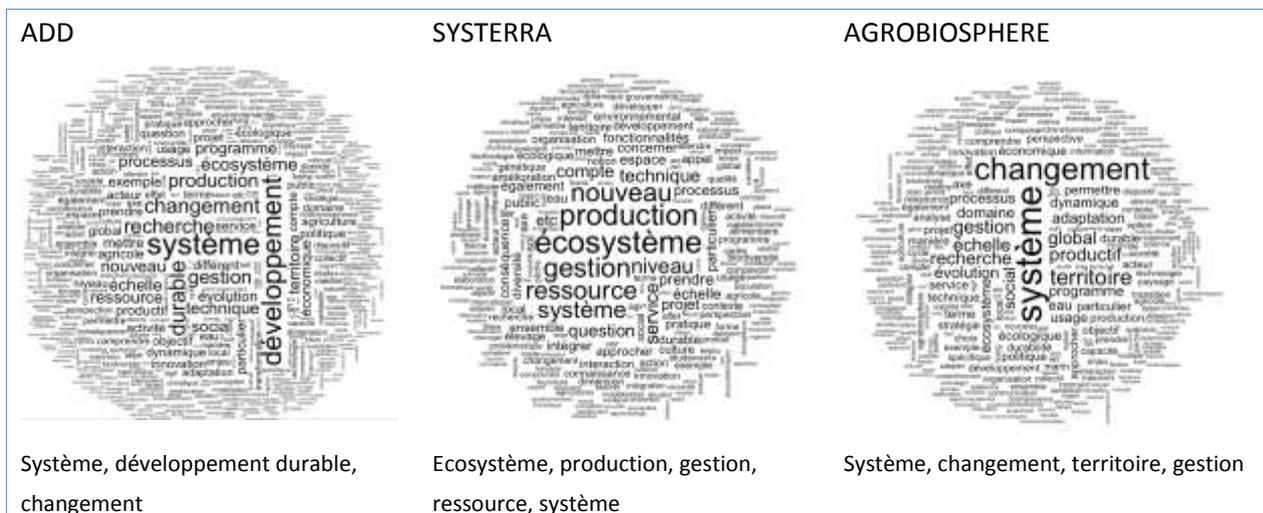
## 2. Résultats

### 2.1. L'offre de programme : les appels comme attracteurs lexicaux

La présente étude est marquée par un choix méthodologique consistant à donner une place centrale aux textes et aux conditions de leur production. Du coup, il n'est pas inutile de caractériser au préalable les trois programmes présentés aux communautés de recherche à partir d'une approche simple de statistique textuelle. Cela peut alors permettre de renvoyer la lecture de la dynamique des projets à la façon dont les trois programmes sont différemment incitatifs. Pour caractériser assez simplement les positions relative des programmes sur le plan lexical, nous avons conduit une Classification Hiérarchique Ascendante avec le logiciel en ligne Iramuteq, avec l'intérêt que présente cette méthode pour repérer les grandes oppositions qui classent un texte ou un corpus de textes. Nous avons également utilisé les analyses proposées comme les nuages de mots et, bien sûr, l'analyse factorielle du nuage des termes extraits et lemmatisés.

Nous avons analysé les textes des trois programmes de la sorte. Une lecture aisée est fournie par les « nuages de mots » qui permettent de représenter ce qu'il en serait d'une lecture globale de ces textes en ne retenant que l'effet de la redondance des mots principaux (Figure 2).

**Figure 2. Nuages de mots des trois textes d'appel à projet – réalisé avec Iramuteq**

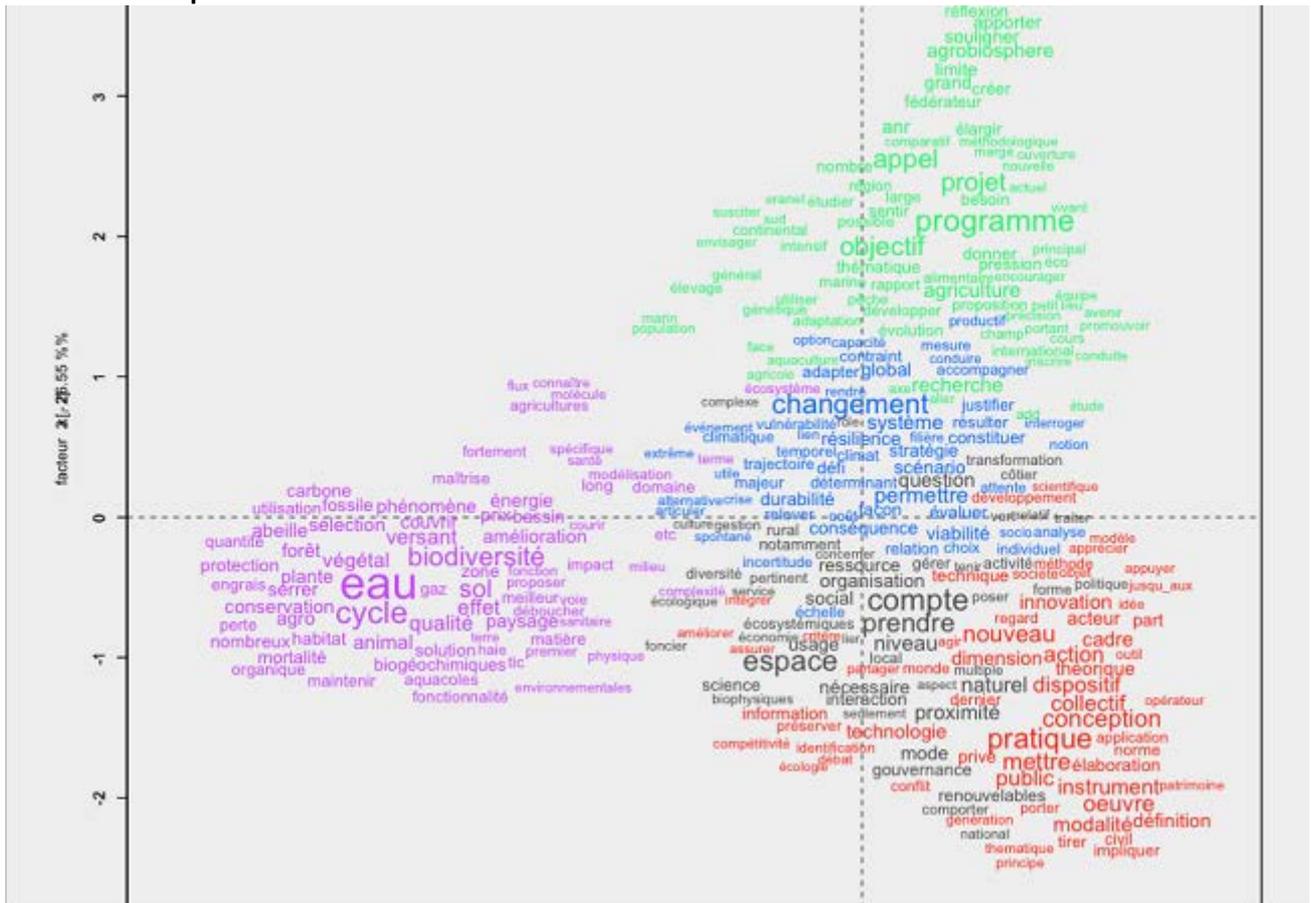


Il est ensuite pertinent de traiter ensemble les textes des trois programmes, tout en conservant le « programme » comme variable. On peut rendre compte des classes de termes obtenus avec une analyse de classification simple. Celle-ci aboutit à 5 classes significatives (sachant que plus de 80% des unités lexicales sont analysées, ce qui est un très bon score) et que 2 axes factoriels expliquent 60% du nuage (voir Figure 3a). Le premier plan factoriel peut faire l'objet d'une description classique.

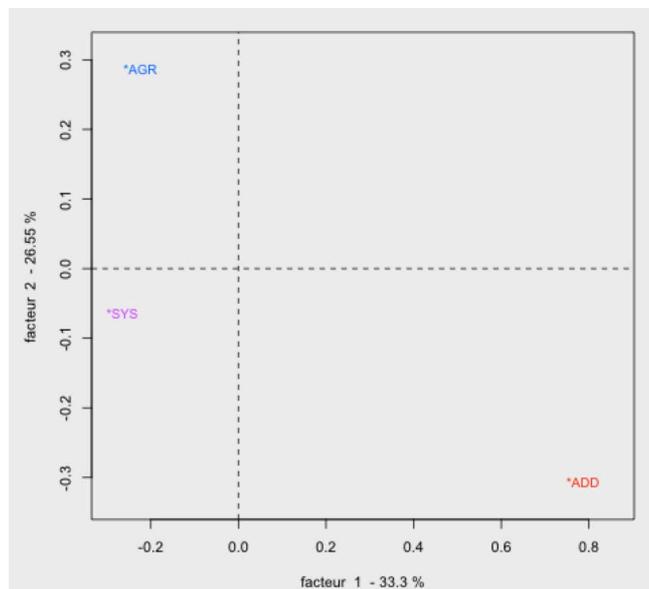
Une thématique (violette) orientée « Ressources » (eau, biodiversité, cycle) apparaît significativement corrélée à SYSTERRA, une thématique (rouge) orientée « Pratiques humaines » significativement corrélée à ADD, une thématique (bleue) orientée « Changement de système » (changement, système, scénario) significativement corrélée à AGROBIOSPHERE. Deux autres classes non corrélées aux programmes apparaissent : une classe (verte) « Discours projet » (programme, objectif, projet) qui identifie une technique locutoire et de positionnement de la recherche dans l'appel ; une classe (noire) « Dispositions » qui identifie des modalités de réalisation de la recherche avec des marqueurs de sa socialisation et de sa

spatialisation. De façon à caractériser des différences sémantiques entre les programmes dans un corpus constitué par les textes des trois appels, nous avons opéré une extraction terminologique des groupes nominaux (500 multi-termes ou N-gram) au seuil de 3 occurrences au niveau de la phrase, puis conduit une MCA sur les 100 plus fréquents en prenant les programmes comme variables. Notre analyse confirme celle obtenue avec de la classification mono-termes (Figure 3b) : il existe une spéciation lexico-sémantique de ces programmes (Figure 4).

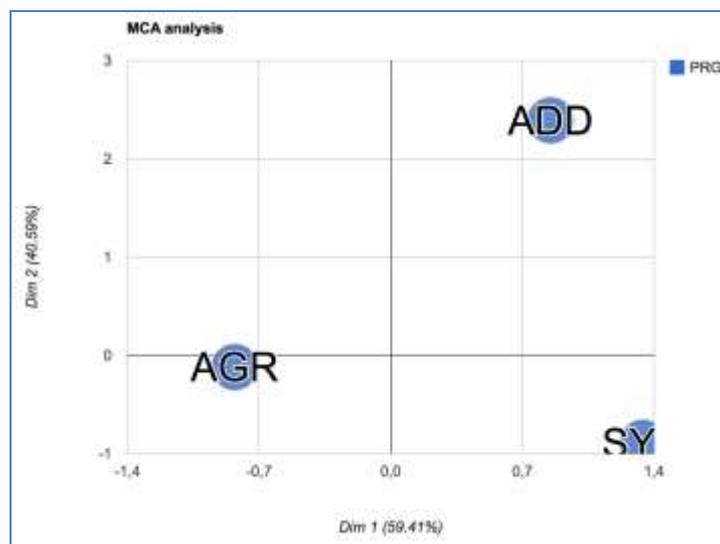
**Figure 3a : Diagramme de l'analyse factorielle des textes des appels à projets (2 premiers axes)- réalisé avec Iramuteq**



**Figure 3b : Premier plan factoriel de l'analyse – projection des variables sur les 2 axes principaux**



**Figure 4 : Résultat MCA : Top 100 N-Gram et Identifiant des Programmes**



*Analyse MCA avec le Module R dans CorTexT-Manager*

De façon à mieux caractériser cette spéciation lexico-sémantique nous avons extrait les clusters des multi-termes qui sont significativement (seuil de Chi-2) attachés du fait de leur fréquence d'occurrence dans les textes des appels (voir les résultats dans l'annexe 2). A partir de cette analyse nous avons opéré une typologie intuitive sur ces listes suivant des catégories indicatrices d'une évolution du positionnement des appels (figure 5) : rapport à la technique, rapport à la production, rapport aux acteurs, rapport au politique et marqueur des grands enjeux du changement.

On tire de cette analyse préliminaire un certain nombre d'enseignements généraux :

- On note que les 3 programmes ont des orientations marquées qui font valoir la notion de système avec des accentuations spécifiques : développement durable, gestion des écosystème et des ressources puis changement au niveau des territoires et de la gestion ;
- Les programmes partagent le souci de valoriser les modalités de conduite et de réalisation de la recherche avec les marqueurs essentiels de sa socialisation et de sa spatialisation ;
- Ils partagent un champ lexical commun autour des modes de production de connaissances attendues.

Les propositions de cadre général que forment ces programmes pour les chercheurs sont suffisamment différentes pour penser que les projets soumis ne s'adressent pas à 3 guichets ouverts sur la même thématique, même si l'orientation générale cible un ensemble de communautés scientifiques de la recherche agronomiques (*sensu largo*) et sur les écosystèmes. Par ailleurs ces appels sont en lien avec l'agenda politique plus général, on peut penser que ADD s'inscrit dans une démarche de création de l'agenda (mention du rapport de O. Godard et B. Hubert), puis SYSTERRA fait écho aux résultats du Grenelle, tandis que AGROBIOSPHERE correspond au moment de la formation du plan ECOPHYTO.

**Figure 5 : Typologie des multi-termes spécifiques des textes de programme**

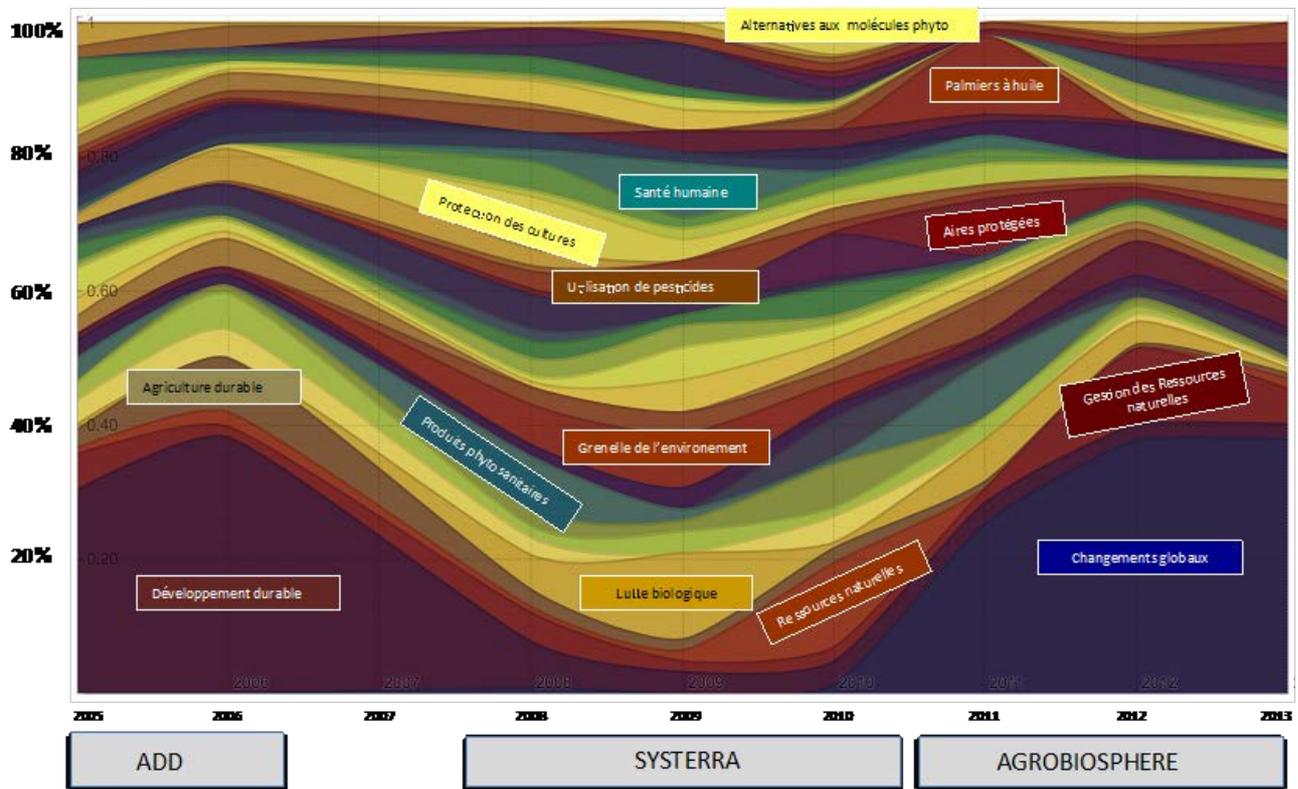
<b>ADD</b>	<b>SYSTERRA</b>	<b>AGROBIOSPHERE</b>
Rapport à la technique		
modèles techniques	itinéraires techniques	choix des cultures
innovation technique	nouvelles techniques	actes de gestion
activités de production	nouvelles technologies	technologies de l'information
processus d'innovation		
Rapport à la production		
entités productives	systèmes productifs	bassins de production
activités agricoles	fonctionnalités des écosystèmes	dynamique des territoires
nouvelles pratiques	services des écosystèmes	couverts végétaux
espace rural	ressources naturelles renouvelables	
Rapport aux acteurs		
acteurs économiques	activités humaines	action collective
agriculture dans le développement durable	collectivités territoriales	stratégies individuelles
différents acteurs	modes de gouvernance	stratégies d'adaptation
générations futures		
Rapport au politique		
politiques publiques	action publique	pouvoirs publics
viabilité économique	biens publics	biens et services
Marqueur des grands enjeux		
gaz à effet	utilisation des produits phytosanitaires	cycle de l'eau

## 2.2. Analyse structurale du paysage sémantique des projets soumis

### 2.2.1. Démographie des multi-termes principaux

Une première analyse de cette population de projets vise à simplement caractériser la démographie des termes extraits des champs textuels riches produit par les auteurs des projets (soient les champs : titre, résumé, mots clés). Par démographie on entend ici un comptage des occurrences des termes extraits et leur proportion pour chaque année de soumission. La Figure 4 propose ainsi une lecture chronologique pour les 25 multi-termes les plus fréquents dans tout le corpus des projets soumis et donnent le pourcentage de projet par année qui contiennent chaque multi-terme retenu. C'est un seuil assez élevé mais il permet de donner un aperçu en restituant visuellement l'idée d'une démographie des termes principaux.

Figure 4. Démographie des Multi-termes extraits : répartition proportionnelle de ces termes



L'interprétation du graphique indique des changements importants : les projets soumis à ADD sont marqués par la thématique du changement durable en agriculture là où ceux soumis à AGROBIOSPHERE consacrent le traitement des enjeux de gestion des ressources dans le traitement de changements globaux. Les projets soumis à SYSTERRA sont clairement contextualisés par le Grenelle de l'Environnement et le renouveau de la protection des cultures induit par les objectifs du Grenelle. Ce dernier ensemble de projets montre une distribution différente de multi-termes et une présence forte de la mise à l'agenda de la réduction des pesticides. Au regard du traitement lexical réalisé sur les programmes, il est à noter que le programme SYSTERRA ne met pas en avant de façon aussi accentuée les objectifs du Grenelle (traité par ailleurs par la partie recherche du Plan Ecophyto), pour autant les projets soumis sont marqués fortement par cet agenda. C'est un indicateur fort du fait que les programmes ont des effets performatifs mais que les communautés de recherche ont aussi une autonomie de positionnement.

scientifique de leurs objets et finalités, en lien ici avec une actualité politique qui polarise l'attention des chercheurs. De plus, il apparaît qu'elles sont en phase avec le traitement des enjeux de société traduit dans l'action publique, puisque le programme Pour et Sur le Plan Ecophyto (PSPE) est en phase de constitution.

### 2.2.2 Le paysage sémantique global des multi-termes

Muni de ces indications, il est possible de caractériser un paysage sémantique des projets soumis en analysant la cooccurrence des multi-termes extraits dans les résumés des projets à partir d'un graphe de réseaux de mots associés. De la sorte la lecture de ces graphes permet d'entrer plus en profondeur dans la composition des associations de termes. Sur cette base, l'analyse restituée dans la Figure 5 donne une indication supplémentaire avec le calcul des clusters de ces termes, entendus comme des regroupements significatifs d'appariements de multi-termes dans les résumés des projets (ces clusters sont indiqués par la couleur des nœuds)<sup>1</sup>. Une lecture macroscopique de ce paysage donne une indication de la partition du travail scientifique sur plusieurs plans thématiques : un ensemble de 8 clusters sur la partie haute du graphe sont proches, au centre 5 clusters sont également proches, tandis que 2 autres sont importants en tailles mais pour ainsi dire peu reliés aux autres (cluster « Changements globaux et adaptation » et cluster « production animale, sélection et bien-être »).

Cette partition reflète des champs sémantiques marqués - en première approche - pour les 8 clusters en haut par des thématiques des sciences du vivant, des sciences économiques et de gestion, et pour les 5 clusters du centre par des thématiques des sciences sociales.

#### Clusters de la partie haute

- 1) Vert gauche: gestion collective durable de ressources/usages, approches pluridisciplinaires
- 2) Violet gauche: écologie des paysages et santé des plantes, génétique et écologie
- 3) Orange centre : économie et décision dans la gestion de ressources naturelles, approche disciplinaire en économie
- 4) Rouge haut: agroécologie du Grenelle et santé des plantes, approche SPE orienté réponse aux enjeux productifs
- 5) Jaune citron; fonctionnement des agroécosystèmes pour la durabilité et le traitement des enjeux; approche pluridisciplinaire centrée agronomie
- 6) Vert droite: agrobiodiversité des milieux et changement/gestion de l'environnement N/S, écologie, pédologie, foresterie
- 7) Violet Droit: modélisation et mesures des cycles biogéochimiques et GES, pédologie biologie
- 8) Bleue: amélioration génétique durable, génétique

#### Clusters de la partie centrale

- 1) Bleu gauche: développement durable, gouvernance, appliquée à différents sujets (HVE, pesticides, intensification écologique, santé animal)
- 2) Vert centre: action collective, acteurs des territoires, analyse compare, dispositif de recherché appliqué à des question de gestion et d'exploitation de ressources et notamment l'eau

---

<sup>1</sup> Rappelons qu'il est issu d'un calcul des clusters via un algorithme de détection de communautés et d'un affichage optimisant leur positionnement relatif, cet aspect rend l'interprétation visuelle de ces graphes aussi robuste que ce que peut l'être toute représentation graphique dans un article scientifique.

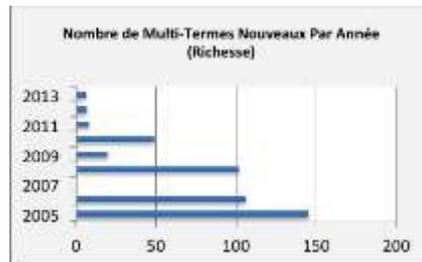






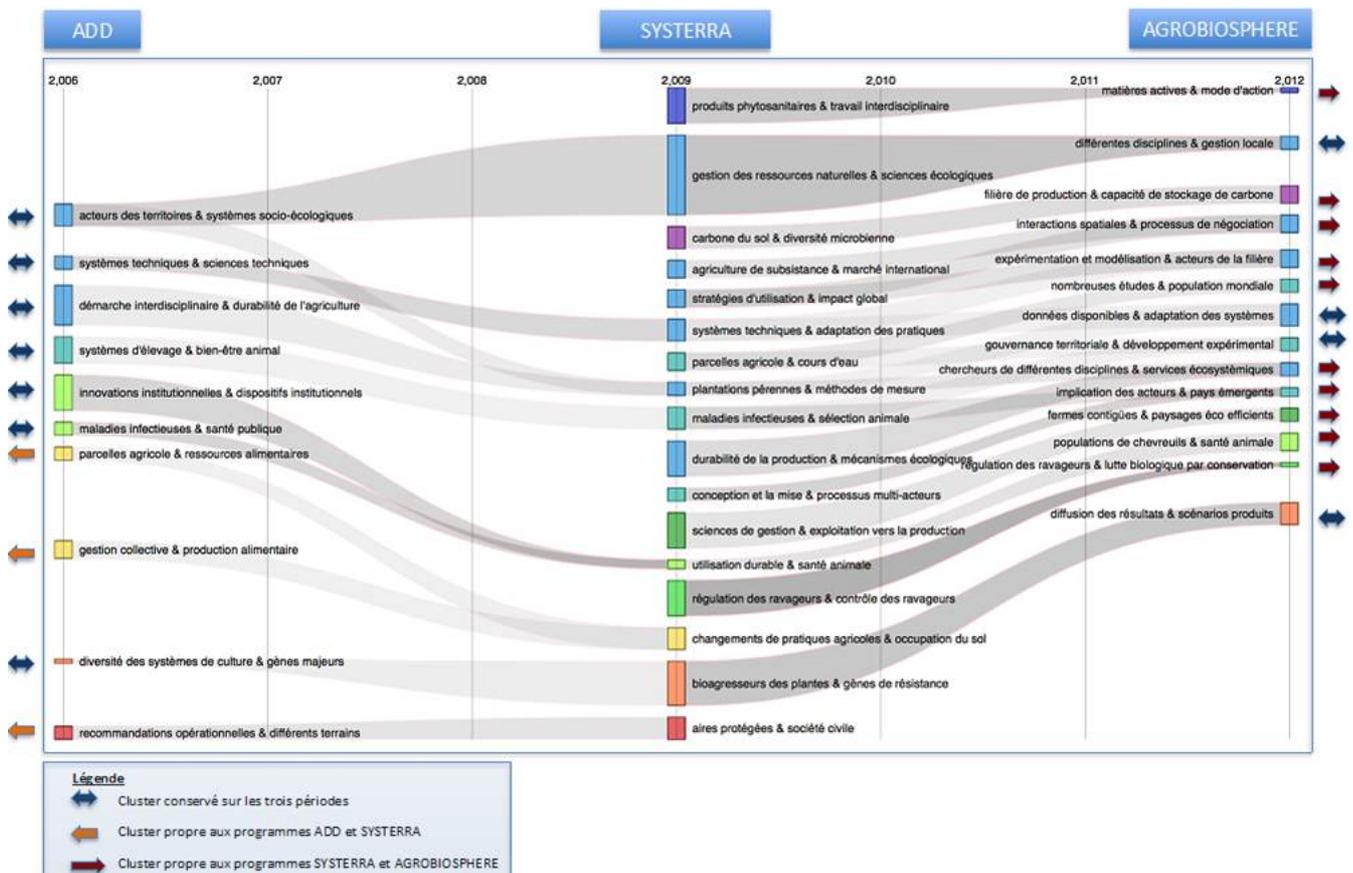
d'approfondissement dans un périmètre lexical déjà fixé. Les porteurs des programmes confirment d'ailleurs ce point. Les réponses faites par les communautés de recherche semblent donc faire écho à la polarisation de l'intention exprimée dans les programmes. C'est un signe de l'effectivité de ce mode de pilotage.

**Graphique 1. Evolution de la richesse lexicale (nouveaux termes entrants).**



Une deuxième voie de caractérisation de la dynamique sémantique consiste à opérer une clusterisation temporelle, c'est à dire en faisant dépendre la construction des clusters à une date T de la structuration des associations existant à la date T-1, toujours suivant l'algorithme de détection de communautés utilisé<sup>3</sup>. Cette relation de dépendance temporelle des clusters d'une année T peut ainsi se manifester par des bifurcations ou à des fusions de clusters d'une année sur l'autre, mais des clusters peuvent apparaître ou au contraire disparaître. C'est une telle représentation de la dynamique sémantique que nous proposons dans la figure 8.

**Figure 8: Carte dite en « tubes » de la composition et de l'évolution temporelle des clusters de termes**



<sup>3</sup> C'est un point fort qu'apporte la plateforme CorTexT et que l'on doit au travail remarquable de Cointet (2009).

Elle fait apparaître les clusters spécifiques à chaque programme et les relations temporelles qu'ils entretiennent successivement. La compréhension de cette figure mérite quelques précisions : la dénomination de chaque cluster est étiquetée par les deux termes les plus importants ; le choix de la palette de couleurs contient une information, les tons indiquant des clusters proches, du fait qu'ils partagent des termes ; le positionnement relatif des clusters au sein de chaque colonne temporelle est optimisé pour permettre une lecture aisée et mettre en évidence les continuités, les chiasmes et les ruptures ; les « tubes » portés en grisé pour relier les clusters d'une année sur l'autre forment une représentation fondée sur un calcul du nombre de termes partagés par les deux clusters, l'intensité mesure visuellement l'intensité de cette relation inter-temporelle. Muni de cette grille de lecture, que nous dit cette carte dite « en Tube » ?

Pour commencer, rappelons nous que nous décrivons les propositions émanant d'un ensemble d'équipe de recherche qui se fédèrent sur des propositions de recherche. Notre interprétation porte sur l'offre de recherche et son évolution. Elle est fondée ici sur le repérage d'une dynamique temporelle dans l'association des multi-termes (en fait des groupes nominaux) dont les agencements fréquents dans les résumés représentent des univers lexicaux pour lesquels nous inférons qu'ils sont porteurs de sens. Le travail interprétatif consiste à les décrire comme des déclarations d'objectifs, d'objets et moyens particuliers de recherche, sorte d'agencement formant des promesses socio-techniques de recherche. Il consiste à relever les continuités (1) et les ruptures (2) de ces agencements.

### **(1) Continuités**

Tout d'abord il s'agit d'appréhender ce graphe en « Tubes » à un niveau de lecture global de l'impact des appels des trois programmes. Il existe des clusters continuent présents sur la durée des trois programmes avec des intensités variables (mesurées par le nombre de projets qui contribuent à la constitution de ces clusters) nous les avons indiqués avec le symbole d'une double flèche dans la figure 8.

- Pris ensemble ces clusters décrivent donc un ensemble de thématiques qui sont indicatrices d'une structuration de la rencontre des incitations portées par les programmes et des promesses de recherche. Pour autant cette permanence dénote un déplacement de positionnements scientifiques volontiers systémiques et en prise sur le développement durable à des positionnements plus spécifiques sur des objets permettant l'étude localisée de la déclinaison d'enjeu énoncé dans le cadre du changement global. Autre point important la revendication d'une interdisciplinarité opératoire semble s'effacer au profit d'un de considérations sur les méthodes, les données et l'implication des acteurs.

- Par rapport à une logique d'exploration sur la contribution de l'agriculture au développement durable que portent les projets soumis au programme ADD, ceux du programme SYSTERRA semblent plus nettement en travail sur la fabrication des constituants d'un régime de recherche finalisé, avec l'importance de la référence aux méthodes, aux données, et le fait que les mondes de l'action concernés par ces recherches sont particulièrement présents.

Pour fonder cette lecture de structuration d'un régime de recherche qui serait engendré par ces trois programmes successifs, le tableau 1 ci dessous donne la liste des clusters permanents qui peuvent représenter, pris ensemble, le cœur scientifique d'un potentiel de recherche associé à cette programmation thématique de l'ANR. Présentés ici sous forme d'intitulés, les thématiques auxquelles renvoient ces intitulés peuvent être mieux décrites par l'ensemble des termes constitutifs des clusters.

**Tableau 1 : Intitulés des clusters dont la composition est permanente dans le temps**

ADD	SYSTERRA	AGROBIOSPHERE
systèmes socio-écologiques § acteurs des territoires	gestion des ressources naturelles § sciences écologiques	différentes disciplines § gestion locales
démarches interdisciplinaires § durabilité de l'agriculture	plantations pérennes § méthodes de mesure	données disponibles – données disponibles
systèmes techniques § adaptation des pratiques	expérimentation modélisation § stratégie d'utilisation et impact global § acteurs des filières	
systèmes d'élevage § bien être animal	maladie infectieuse § sélection animale	implications des acteurs § pays émergents
innovations institutionnelles § dispositifs institutionnels	utilisation durable § santé animale	population de chevreuils § santé animale
diversité des systèmes de cultures § gènes majeurs	Bio-agresseurs des plantes § gènes de résistance	diffusion des résultats § scénario produits

**(2) Discontinuités**

Ensuite, il s'agit d'appréhender la carte de la figure 8 dans le but d'identifier des discontinuités, à savoir des ruptures et des émergences dans l'agencement induit à la croisée des incitations programmatiques et des promesses de recherche. Des clusters apparaissant dans ADD sont maintenus dans SYSTERRA, mais disparaissent dans AGROBIOSPHERE : ils sont regroupés dans le Tableau 2a ci-dessous.

D'autres n'existent pas au départ, mais émergent dans SYSTERRA et sont maintenus dans AGROBIOSPHERE. Dans ce deuxième mouvement on peut noter des structurations émergentes stables (Tableau 2b *stables*), et celles plus nombreuses qui procèdent de recompositions de clusters par fusion ou dissociation, et qui sont plus difficile à interpréter, notamment parce que le nombre de projets soumis à AGROBIOSPHERE est moins important. De manière générale, tout se passe comme si la variété créée par SYSTERRA tendaient à se recomposer dans des clusters plus spécialisés et moins nombreux dans AGROBIOSPHERE.

**Tableau 2a.**

ADD	SYSTERRA
recommandations opérationnelles § différents terrains	aires protégées § société civile
gestion collective § production alimentaire (Fusion) parcelles agricoles § ressources alimentaires	changements de pratiques agricoles § utilisation du sol

**Tableau 2b. Stables**

SYSTERRA	AGROBIOSPHERE
produits phytosanitaires § travail interdisciplinaire	matières actives § mode d'action
régulation des ravageurs § contrôle des ravageurs	régulation des ravageurs § lutte biologique par conservation
carbone du sol § diversité microbienne	filiale de production – capacité de stockage du sol
sciences de gestion § exploitation vers la production	fermes contiguës § paysages éco-efficients



## 2.4. Le paysage des collaborations entre unités de recherche

### 2.4.1. Le sens des assemblages

On ne peut aborder les aspects sémantiques de ces textes sans considérer que cette littéracie scientifique des projets a des auteurs.

La rédaction d'un projet est une épreuve collective, dont la conduite est portée par un responsable de projet (parfois 2) qui assume une fonction éditoriale. L'implication des chercheurs dans ces projets procède de multiples façons et elle varie en intensité. C'est une activité de production scientifique beaucoup moins connue que la production d'article scientifique qui a connu un investissement certain depuis les premiers travaux de Gusfield (1976), Latour et Fabbri (1977) et O'Neill (1981), par exemple. L'écriture du texte scientifique doit être envisagée comme un exercice rhétorique, pour lequel les auteurs puisent dans la scène du laboratoire et dans les connaissances certifiées du domaine, les ressources pour convaincre les lecteurs et augmenter son capital dans le champ scientifique (Latour et Fabri, 1977). Dans le cas de l'écriture des projets, la scène du laboratoire n'est pas le seul espace de référence. Si on considère le projet comme une « œuvre littéraire », elle correspond à une écriture à plusieurs mains faite de négociations et d'un arbitrage final (souvent dans l'urgence ...) qui fait du responsable de projet un genre d'éditeur, sa signature recouvrant des signatures multiples. En considérant ces projets comme des textes scientifiques à part entière, on fonde une approche intéressante pour en faire des traces de dynamiques sociales et professionnelles sous-jacentes et en cours d'établissement. Le projet n'inscrit pas l'ensemble de cette dynamique, et notamment des négociations qui ont présidé à la mise en évidence de tel ou tel argument. Une étude sociale de cette écriture serait à fournir, mais elle n'était pas dans nos objectifs avec cette étude. Qui plus est, afin de ne pas personnifier notre regard sur ces textes, nous allons nous en tenir à l'analyse des relations entre les laboratoires, marqueur de l'affectation des auteurs de ces projets.

Nous garderons donc à l'esprit cette façon de concevoir le projet comme œuvre collective éditée et co-signée, pour aborder le traitement d'une information importante produite grâce au travail de création de la base de données avec ProjBase. En effet les projets sont portés par des chercheurs qui appartiennent à des laboratoires qui ont la durée de vie que leur donnent leurs tutelles. La base de données contient ainsi une information organisationnelle à travers la façon dont les projets mettent ces laboratoires en relation. La répétition de ces associations dans le temps indique des formes de structuration des relations de travail entre les laboratoires. Ainsi la dynamique des promesses portées par les projets est aussi une dynamique sociale et professionnelle. On notera que les informations traitables à ce stade ne permettent pas de « descendre » au niveau des équipes de recherche, mais il est possible d'analyser le niveau des unités participantes aux projets, au sens de recevant une subvention et soumis à la justification de leurs activités. On notera que, parfois, les affiliations sont données au niveau des établissements, ce qui introduit un biais lié aux données, mais ces cas sont plutôt rares en nombre de collaborations, ils ne viennent pas gauchir l'analyse.

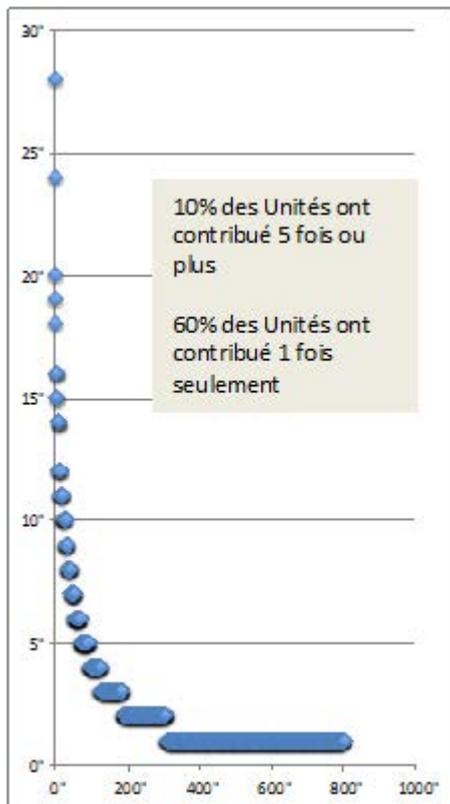
### 2.4.2. Un mouvement d'extension des collaborations

Un premier niveau de résultat est donné par l'existence d'un *pool* d'unités qui ont déposé fortement dans les trois programmes (10% des unités l'ont fait plus de 5 fois) et tout un ensemble d'unités nombreuses (plus de 60% du total) ne sont participantes qu'une seule fois (voire graphique 2).

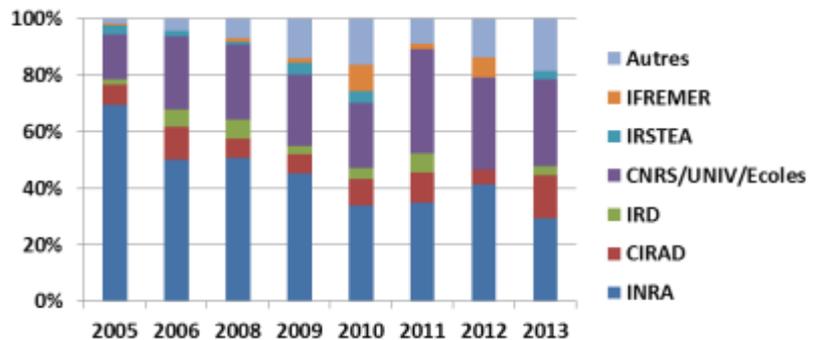
Cette distribution reflète une loi de puissance et elle n'est pas surprenante : des laboratoires de taille importante ont plus de capacité de soumettre du fait du nombre de chercheurs, des laboratoires peu

positionnés dans le domaine peuvent rejoindre un projet du fait de collaborations particulières entre 2 chercheurs ou 2 équipes. Cette distribution indique que les incitations des programmes mettent en visibilité une structuration d'unités qui sont des attracteurs pour d'autres unités, ainsi que parfois pour des acteurs de la recherche appliquée. Ainsi, les données concernant les établissements impliqués dans les collaborations renforce cette idée d'une programmation thématique qui est aussi un dispositif d'incitation et d'exploration des collaborations possibles sur les objectifs scientifiques portés par les projets. Enfin, trait important, on note que sur la durée des ces 3 programmes des établissements de plus en plus variés se sont trouvés impliqués à partir d'une contribution initiale à ADD très forte et structurant de l'INRA (graphique 3).

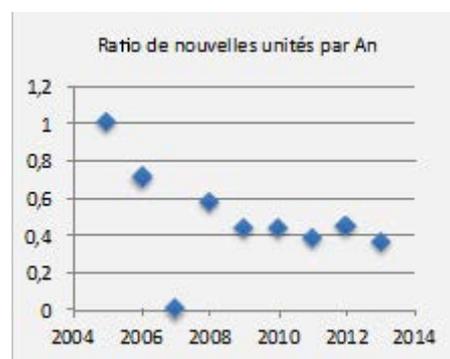
Graphique 2



Graphique 3



Graphique 4



Pour préciser ce mouvement d'extension des collaborations, il est pertinent de faire le ratio année après année du nombre d'unités nouvelles apparues dans l'année N par rapport à celles présentes dans l'année N-1 (graphique 4). L'évolution de ce ratio décroît sur les trois premières années indiquant un possible effet de saturation des collaborations possibles, puis se stabilise à 0,4 sur les 6 années restantes. C'est un indicateur de plus de ce mouvement d'extension qui maintient une capacité d'exploration des collaborations scientifiques. Il milite pour confirmer l'idée d'un régime de recherche en cours de stabilisation, ici, dans sa composante professionnelle.

L'impact connaît ainsi deux phases : la première auprès d'une population spécifique d'unités qui ancre nettement le travail de recherche sur la thématique du développement durable de l'agriculture, puis une phase d'expansion dans une population d'unités qui s'enrichit continuellement. On manque encore de recul pour comparer ce type de distribution des collaborations entre plusieurs programmations thématiques. Mais cela semble un outil pertinent pour qualifier l'impact des programmes sur les communautés de recherche. Voyons maintenant de quoi est fait ce mouvement de structuration des collaborations.

### 2.4.3. La dualité de l'organisation par projets

Progressons dans notre analyse de cette population d'unités impliquées dans l'espace de recherche thématique ouvert par les 3 programmes. Toujours en mobilisant la méthode d'analyse des mots-associés et des traitements dans la plateforme CorText, on peut rendre compte du réseau des associations des unités telles que la participation aux projets l'établit. Une *clusterisation* de ce graphe de tous les projets soumis (donc quelque soit le programme, c'est à dire sans lecture temporelle) permet de caractériser l'espace des collaborations institutionnelles au grain de l'unité, sachant qu'elles sont portées par des implications directes de chercheurs.

La carte de la figure 10 restitue ce calcul du graphe et des clusters : la taille des nœuds est proportionnelle au nombre de fois qu'une unité a été impliquée dans des collaborations pour la contribution à projet et l'épaisseur des liens indiquant la répétition des relations ; les cercles colorés indiquent quant à eux les relations significativement importantes du fait de la fréquence d'association prise 2 à 2. Sur ce graphique nous avons apporté une information supplémentaire en identifiant les unités fortement contributrices (par ailleurs souvent de taille importante) qui établissent des liens entre ces clusters ; cette information est rajoutée sur la carte initiale livrée par CorText Manager, sous la forme de cercles bleus.

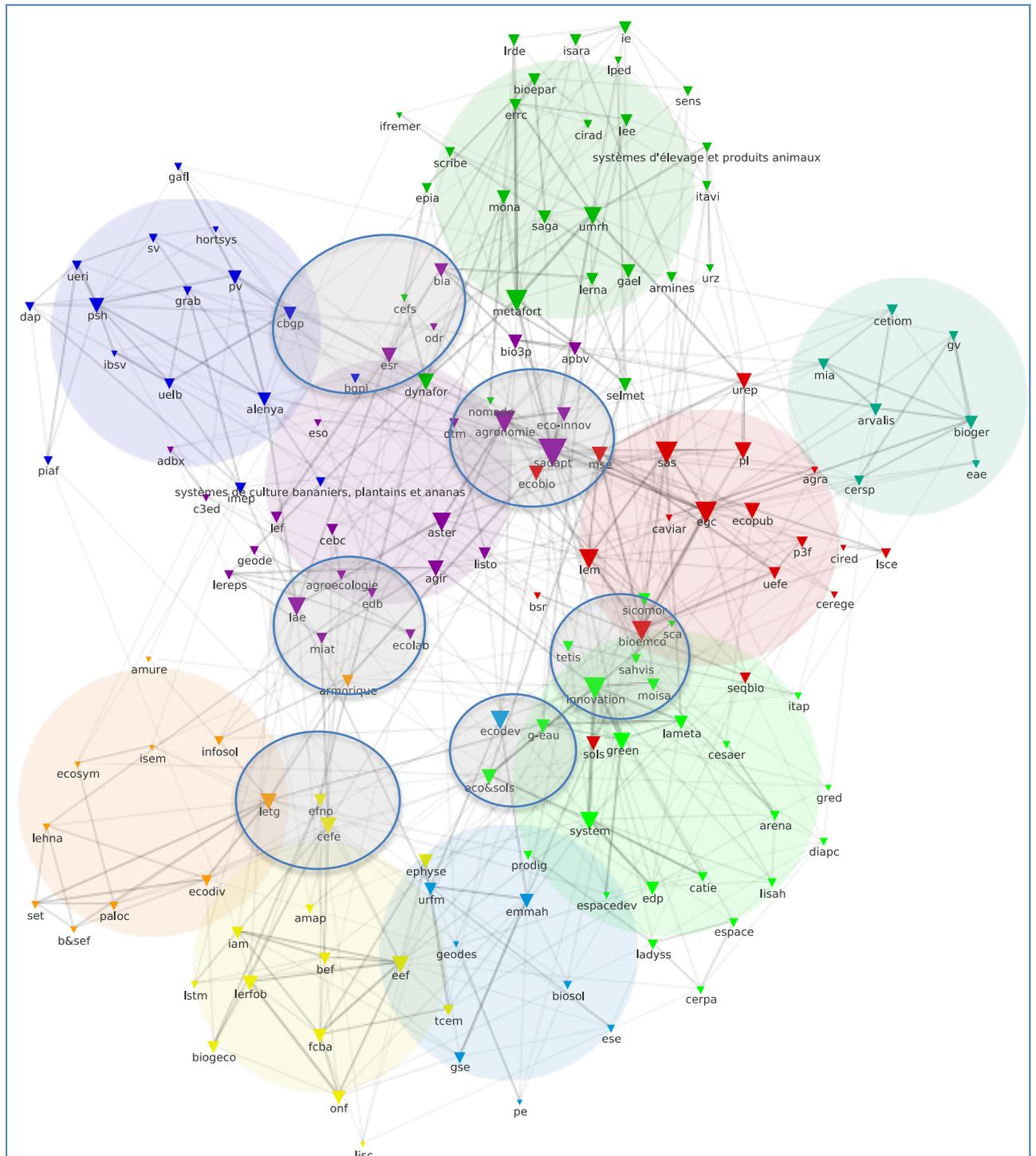
Cette carte permet de repérer les unités qui sont les principaux attracteurs du programme par leur niveau de contribution aux collaborations (la taille des triangles indique le nombre de liens du nœud) et de noter qu'elles ont une position souvent d'intermédiaire ou de carrefour de collaborations, même si elles contribuent de façon significative pour construire un cluster de collaborations particulières. Les collaborations intra-cluster peuvent être interprétées comme relevant d'une logique de « club », reflétant des réseaux de collaborations significativement forte entre des équipes en position d'équivalence structurelle dans le réseau du cluster ; les positions de certains laboratoires entre des clusters sont elles indicatrices d'un type de collaborations différentes (même si la taille des nœuds est faible ou si leur position est en dehors du cluster<sup>5</sup>), leur position intermédiaire traduit une position de carrefour.

La topologie de ces clusters indique une structuration des collaborations autour de trois clusters plus centraux : avec le cluster violet dont SADAPT et AGRONOMIE sont les attracteurs importants, le cluster rouge dont EGC et SAS forment les attracteurs majeur en lien avec BIOEMCO LEM et LIREP qui sont des composantes plus périphériques mais déterminantes de relations externes à ce cluster ; le cluster vert tendre dont INNOVATION, SYSTEM, LAMETA et GREEN sont les attracteurs majeurs. Dans la partie basse de la carte, certaines unités ont des positions notables de carrefours de collaborations comme ECODEV, ARMORIQUE, LAE, LETG, CEFE. On retrouve aussi dans la partie haute des unités de ce type avec METAFORT, DYNAFOR, BIO3P, SELMET, APBV, UREP. Gardons à l'esprit que nous considérons ici les projets soumis, et que des laboratoires non lauréats peuvent resoumettre d'un appel à l'autre.

---

<sup>5</sup> Précisons que ces cercles donnent par leur diamètre une information sur le nombre de projets concernés mais ils ne contraignent pas les relations entre les nœuds, il s'en suit que des nœuds peuvent se trouver à l'extérieur des cercles

Figure 10. Carte des collaborations entre les unités recevant une subvention



Ce paysage des collaborations est donc indicateur d'une valence forte de ces laboratoires structurants du paysage pour les thématiques des programmes, et d'une certaine spécialisation de leurs promesses scientifiques qui s'accompagne d'une capacité à relier des laboratoires moins centraux dans leurs soumissions. On peut interpréter cette topologie des projets comme un impact de la programmation thématique de l'ANR : d'une part rendre visible, légitimes des laboratoires centraux ou carrefours (ceux qui sont attracteurs dans les clusters et qui sont porteurs des relations entre les clusters) d'une part, et rendre possible des ensembles de collaborations avec ces laboratoires suivant une logique de club thématique (les relations de collaborations intra-cluster).

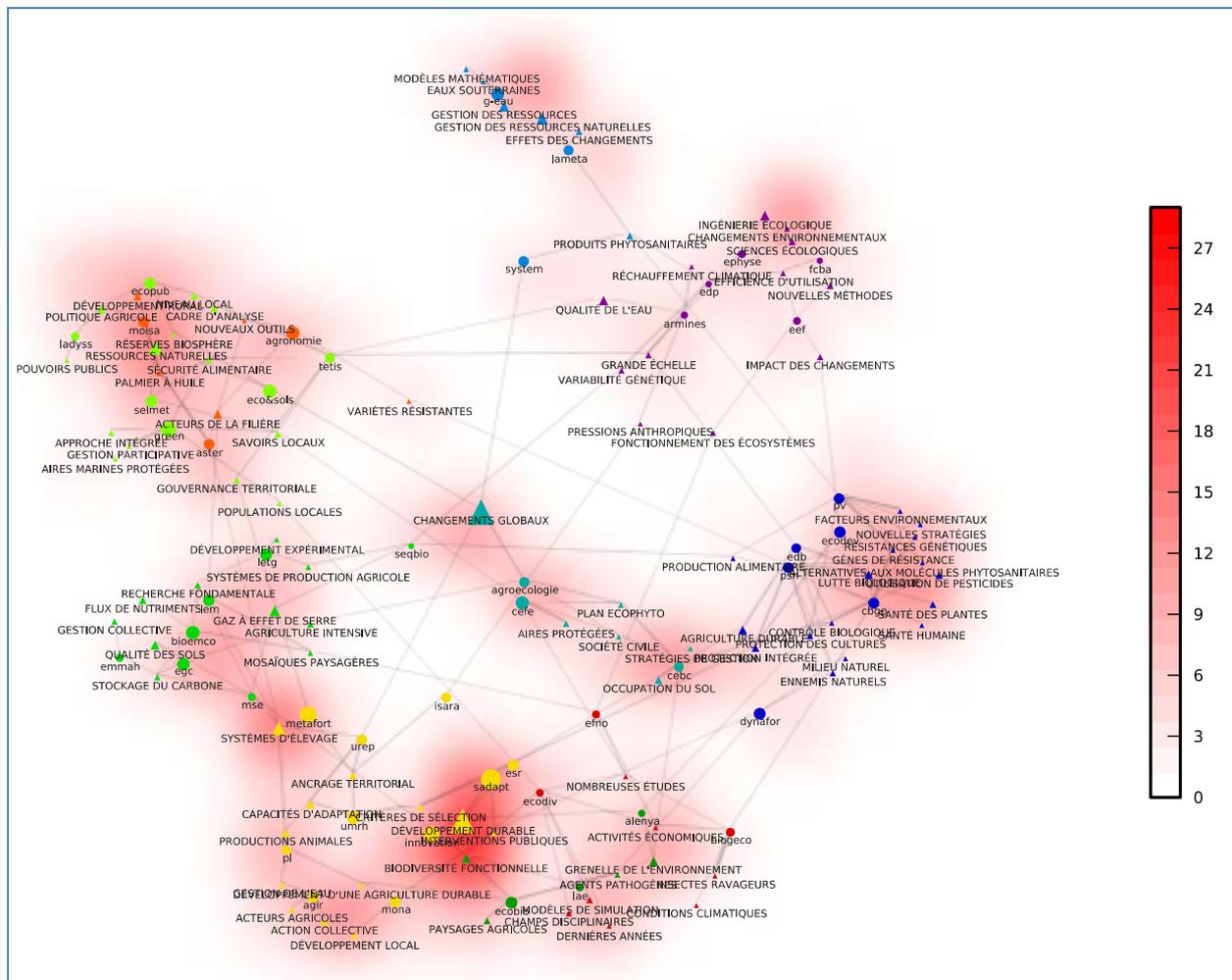
Ce travail d'identification pourrait être alors complétée par une description de ce que représente ces unités du point de vue disciplinaire ou interdisciplinaire, de leur taille, de leur tutelles etc. Un tel travail suppose une caractérisation des unités sur le plan de leur localisation géographique, de leur profil disciplinaire, voire de leur taille, il reste à conduire à partir d'un travail plus poussée sur les données.

#### 2.4.4. La composition socio-sémantique des collaborations

Pour finir notre description, reste à questionner la façon dont ces trois programmes ont pu favoriser le positionnement, voire la spécialisation, des unités sur des thématiques résultantes de la projection de leur capacités, ressources et promesses scientifiques. Pour instruire cette question, il est pertinent de considérer les associations entre les termes extraits des résumés et les laboratoires.

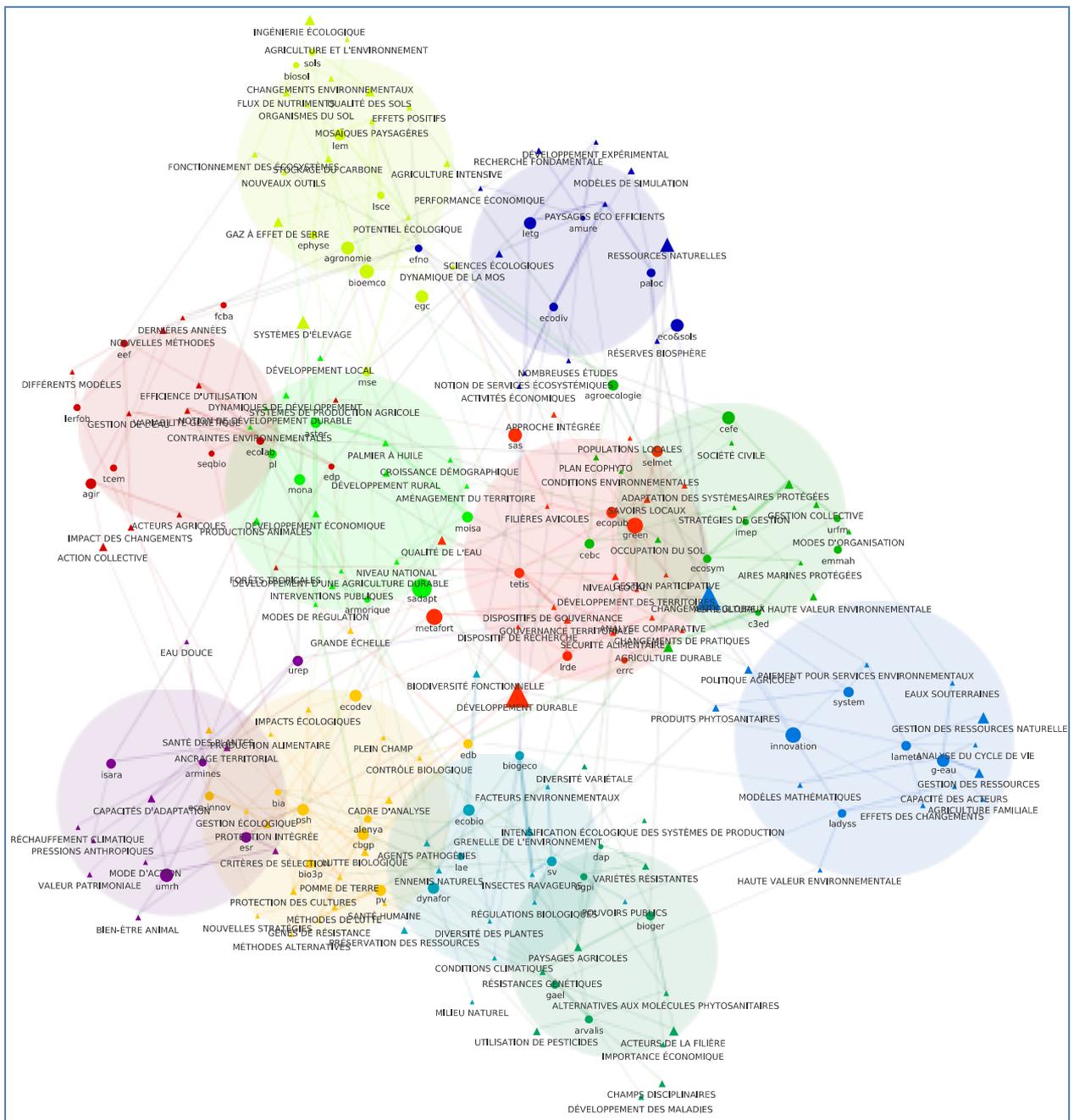
Cela conduit à la production d'un graphe bipartite sur lequel il est possible de projeter la variable « projet financé » pour appréhender comment les décisions de subventions ont impacté ces agencements (voir figure 11). Il ressort qu'à ce niveau de lecture (à savoir en ayant sélectionné les 50 laboratoires les plus impliqués dans les collaborations et les top 100 N-grams), les décisions de financement impactent l'ensemble des associations hétérogènes Labo / N-gram. Un paramétrage plus fin avec 75 laboratoires et 250 termes donne le même résultat. Cela signifie que les décisions de financement n'ont pas produit une sélection particulière sur cette offre de promesses.

**Figure 11. Carte des associations hétérogènes entre laboratoires et Termes avec projection de la variable « projet financé » (heatmap avec intensité du Chi2)**



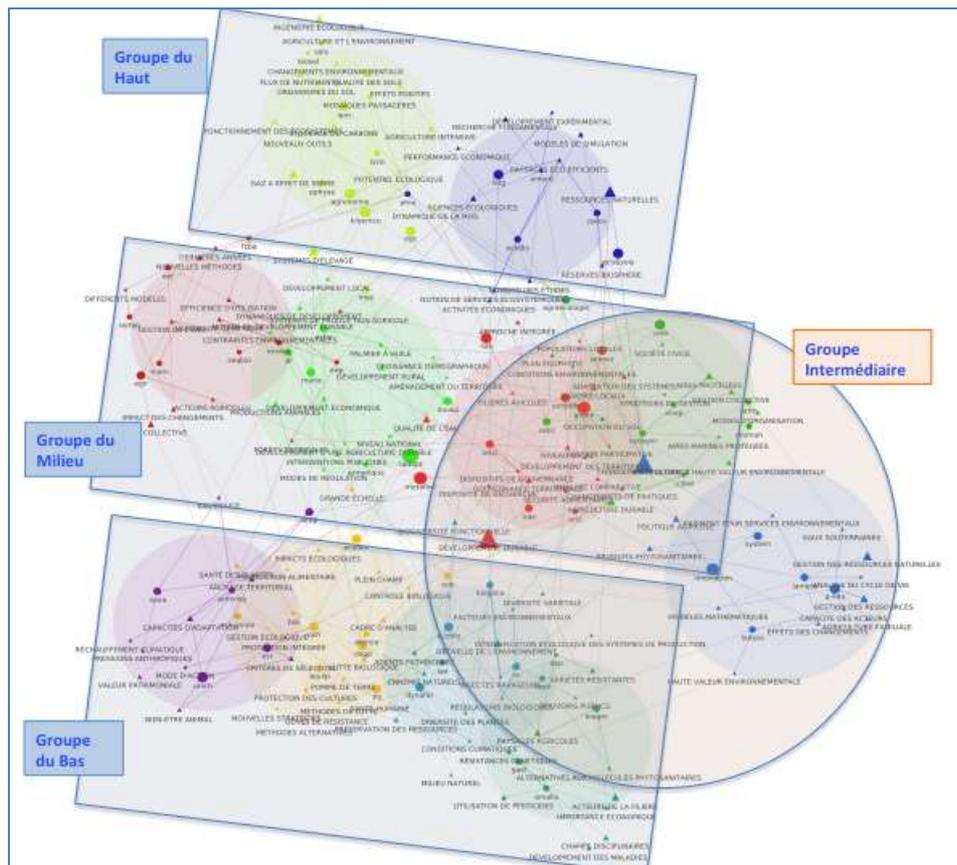
Après avoir « neutralisé » la possibilité d'un biais de sélection sur les agencements hétérogènes de laboratoires et de multi-termes, observons maintenant la composition de ces agencements sur l'ensemble des trois programmes dans le but de caractériser l'impact de cette programmation thématique de l'ANR. Nous avons sélectionné les 150 laboratoires et les 150 multi-termes les plus fréquemment impliqués dans les soumissions ; la métrique est Chi2 sans seuillage sur le nombre de liens ; le calcul de la clusterisation par Louvain est optimisé. Les laboratoires sont figurés par des cercles, et les multi-termes sont figurés par des triangles (la taille du nœud est proportionnelle au nombre de liens du nœud concerné). L'interprétation du graphe proposé dans la figure 12 vient éclairer celui des collaborations.

**Figure 12. Carte des clusters d'associations hétérogènes des laboratoires et multi-termes**



Pour conduire l'interprétation de ce graphe, on renvoie tout d'abord à la lecture la carte dans la figure 13 qui propose une interprétation « macroscopique », à savoir reposant sur une description de la topologie des clusters, en considérant leur disposition relative et les liens forts qu'ils entretiennent entre eux de façon privilégiée. Sur le fond de carte du calcul de la clusterisation (reprise de la figure 12) nous avons porté en surimpression des indications de lecture de cette topologie. A ce niveau de lecture cette structuration des groupes peut se lire comme l'impact de ces programmes sur les communautés de recherche.

**Figure 13. Paysage des groupes de cluster : l'impact des programmes**

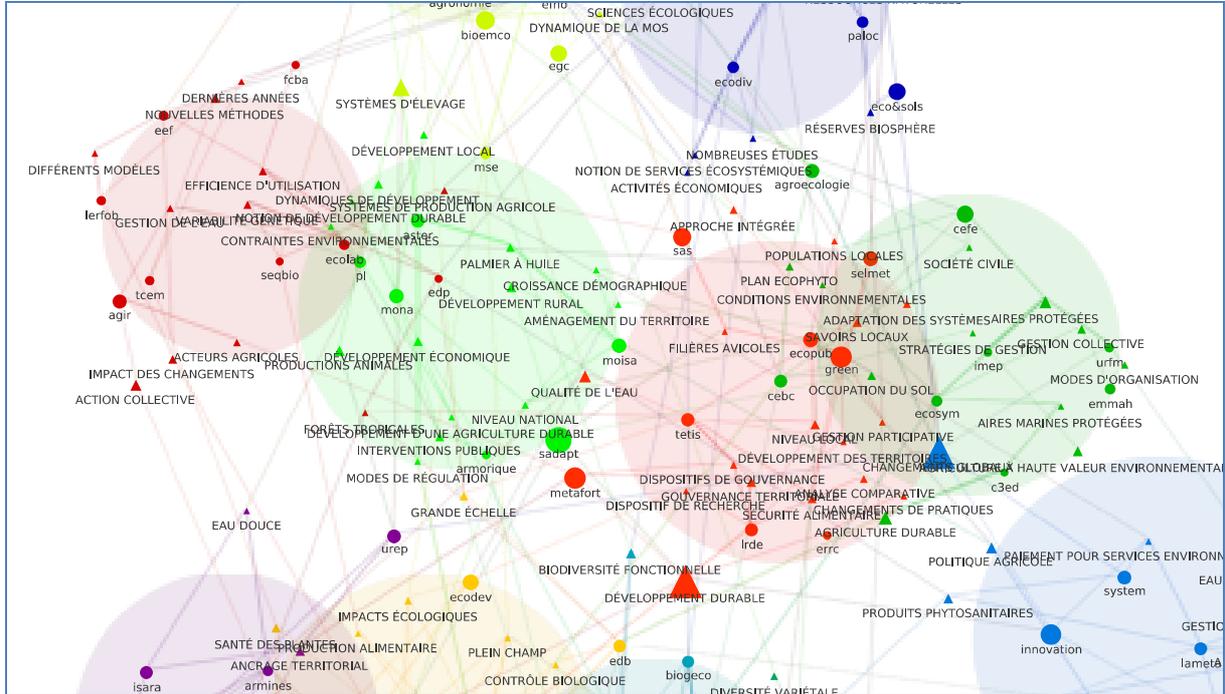


Nous observons ainsi trois niveaux de clusters agencés comme en couches (groupe haut, bas et du milieu), et, latéralement, un cluster important par sa taille (donc relative au nombre de projets concernés) qui se trouve relié à des clusters de deux des trois niveaux (groupe intermédiaire). Chaque groupe peut être ainsi interprété comme un ensemble cohérent de promesses scientifiques portées par des laboratoires, et cela sur la durée des trois programmes successifs. Cette approche de la topologie d'ensemble doit être complétée en abordant de façon plus précise les contenus des clusters et les relations de ces groupes (voir les figures 14 a – d qui suivent). En nommant ces groupes nous portons notre interprétation de ces agencements à travers la formulation de leur étiquetage en prenant sciemment le terme de « sciences » pour signifier l'impact des trois programmes dans la constitution d'un potentiel de recherche (puisque qu'il s'agit des projets soumis) :

- Sciences de l'écologie des sols et des paysages (Gpe du Haut);
- Sciences de la durabilité de l'agriculture par la régulation des activités humaines (Gpe du milieu) ;
- Sciences de la gestion à haute valeur environnementale des ressources (Gpe intermédiaire) ;
- Sciences de la régulation écologique des pressions biotiques et anthropiques (Gpe du Bas).

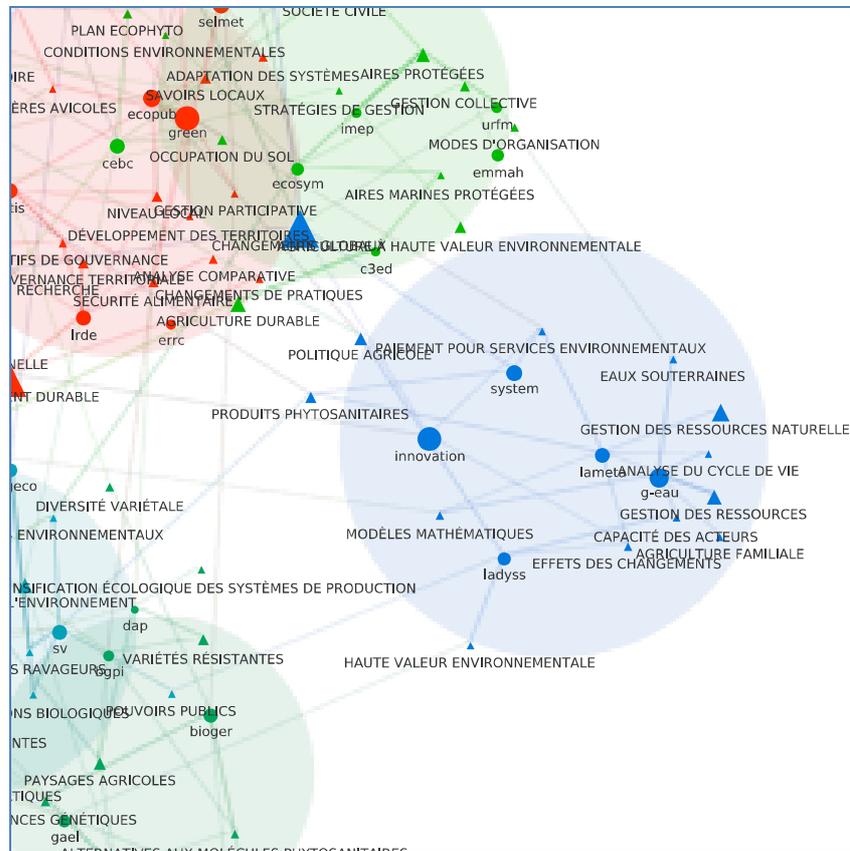


Figure 14 b (Groupe du milieu) : Sciences de la durabilité de l'agriculture par la régulation des activités humaines



<b>ROUGE (Gauche)</b>	<b>VERT (Centre)</b>	<b>ROUGE (Centre)</b>	<b>VERT (Droite)</b>
TCEM	SADAPT	TETIS	CEFE
ECOLAB	PL	SELMET	EMMAH
EEF	MOISA	SAS	IMEP
AGIR	MONA	METAFORT	C3ED
EDP	ASTER	LRDE	CEBC
LERFOB	ARMORIQUE	GREEN	ECOSYM
FCBA		ECOPUB	URFM
SEQBIO		ERRC	AGROECOLOGIE
différents modèles	dynamiques de développement	qualité de l'eau	société civile
forêts tropicales	productions animales	changements de pratiques	Aires Marines Protégées
systèmes de production agricole	développement économique	dispositif de recherche	aires protégées
acteurs agricoles	interventions publiques	approche intégrée	gestion collective
action collective	modes de régulation	savoirs locaux	stratégies de gestion
dernières années	développement rural		agriculture à HVE
variabilité génétique	aménagement du territoire	analyse comparative	plan Ecophyto
nouvelles méthodes	croissance démographique	sécurité alimentaire	conditions environnementales
efficacité d'utilisation	palmier à huile	dispositifs de gouvernance	occupation du sol
gestion de l'eau	développement local	gouvernance territoriale	agriculture durable
impact des changements	contraintes environnementales	adaptation des systèmes	modes d'organisation
	notion de développement durable	filières avicoles	
	niveau national	populations locales	
	développement d'une agriculture durable	développement des territoires	
		gestion participative	
		développement durable	
		niveau local	

**Figure 14 c (groupe intermédiaire) : Sciences de la gestion à haute valeur environnementale des ressources (eau)**



<b>BLEU DROITE</b>
INNOVATION
LADYSS
G-Eau
LAMETA
SYSTEM
gestion des ressources naturelles
gestion des ressources
modèles mathématiques
haute valeur environnementale
politique agricole
effets des changements
capacité des acteurs
paiement pour services environnementaux
eaux souterraines
produits phytosanitaires
agriculture familiale
Changements globaux
analyse du cycle de vie



A la suite de cette lecture du paysage globale sur l'ensemble des appels des trois programmes, il est pertinent de traiter la possibilité que certains des agencements hétérogènes de laboratoires et de multi-termes soient spécifiques à certains appels. Pour étudier cette possibilité nous avons réalisé une série de 3 cartes permettant de porter la projection de la contribution de ces agencements aux trois programmes (figure 15).

Il ressort que le programme ADD a particulièrement impacté la constitution d'une offre de Sciences de la durabilité de l'agriculture par la régulation des activités humaines (Groupe du milieu) ; le programme SYSTERRA marquent de façon significative la constitution d'une offre de Sciences de la régulation écologique des pressions biotiques et anthropiques (Groupe du Haut) et de Sciences de l'écologie des sols et des paysages (Groupe du Haut) ; enfin le programme AGROBIOSPHERE impacte la constitution d'une offre des Sciences de la durabilité de l'agriculture par la régulation des activités humaines (Groupe du milieu). On tire en tous les cas de ce travail de projection des programmes sur les agencements de laboratoire et de mots-clés, l'idée que ces programmes engendrent une logique d'offre et de portage de promesses scientifiques à la fois globale mais aussi suivant certaines intensités particulières. On peut ainsi retrouver ici comment cette spécialisation de l'offre fait écho aux différences que nous avons notées précédemment dans le lexique des trois programmes.



## Conclusion

Nous avons tenté avec ce travail de positionner une orientation de recherche sur l'étude socio-sémantique de bases de données qui résultent du travail d'incitation et de sélection réalisé par des agences de pilotage de la recherche publique. L'objectif était de parvenir à caractériser les effets d'une programmation thématique sur la constitution d'une offre de promesses scientifiques et de leur agencement corrélatifs de laboratoires porteurs. Nous avons tout d'abord proposé de positionner notre travail en insistant sur l'idée que l'écriture du projet de recherche correspondait à une littéracie scientifique encore peu étudiée. La perspective ouverte par le traitement des bases de données de projets ne visait pas la constitution d'une critique a priori du managérialisme qui s'étend dans la recherche, indubitablement. On a surtout voulu caractériser les agencements que suppose cette forme d'écriture scientifique sans replier pour autant nos analyses sur des considérations purement linguistiques, mais plutôt en comprenant ces agencements comme un travail de *self-accountability* de l'organisation de la recherche par les promesses.

Un acquis de cette étude est d'avoir permis de structurer une base de données de projets soumis grâce à une collaboration intense avec l'ANR. Il ressort de ce travail une application disponible, dans une technologie non-proprétaire : projbase. Cette application est un prototype constitué pour répondre tant à des exigences de gestion des données, qu'à des objectifs d'analyse de ces données, notamment par une capacité d'exportation très utile pour le chercheur.

Un résultat préliminaire concerne l'analyse lexicale des appels à projets, travail préalable nécessaire pour comprendre les orientations de cette programmation thématique et caractériser son évolution puisque trois programmes ont été analysés. On tire de cette analyse un certain nombre d'enseignements généraux. Les 3 programmes ont des orientations marquées qui font valoir la notion de système avec des accentuations spécifiques : développement durable, gestion des écosystème et des ressources puis changement au niveau des territoires et de la gestion. Ces programmes partagent le souci de valoriser les modalités de conduite et de réalisation de la recherche avec les marqueurs essentiels de sa socialisation et de sa spatialisation. Ils ont un champ lexical commun autour des modes de production de connaissances attendues. Cependant, les propositions sont suffisamment différentes pour penser que les projets soumis ne s'adressent pas à 3 guichets ouverts sur la même thématique, même si l'orientation générale cible un ensemble de communautés scientifiques de la recherche agronomiques (*sensu largo*). On notera enfin que ces appels sont en lien avec l'agenda politique plus général, on peut penser que ADD s'inscrit dans une démarche de création de l'agenda, puis SYSTERRA fait écho aux résultats du Grenelle, tandis que AGROBIOSPHERE correspond au moment de la formation du plan ECOPHYTO.

Viennent ensuite les résultats importants en réponse à l'objectif de caractérisation des effets d'une programmation thématique. Tout d'abord sur le plan sémantique (1), ensuite sur le plan des collaborations de laboratoires (2), enfin en combinant ces deux niveaux d'analyse (3)

(1) Nous avons tout d'abord conduit un premier niveau de lecture des projets soumis aux appels pour caractériser le paysage sémantique de l'offre de recherche globale qui en résulte à travers la démographie des multi-termes extraits sur l'ensemble du corpus formé par les résumés des projets. Nous mettons en évidence que les projets soumis à ADD sont marqués par la thématique du changement durable en agriculture là où ceux soumis à AGROBIOSPHERE consacrent le traitement des enjeux de gestion des ressources dans le traitement de changements globaux. Les projets soumis à SYSTERRA sont nettement

contextualisés par le Grenelle de l'Environnement et le renouveau de la protection des cultures. Ce dernier ensemble de projets montre une distribution différente de multi-termes et une présence forte de la mise à l'agenda de la réduction des pesticides. Au regard du traitement lexical réalisé sur les programmes, il est à noter que le programme SYSTERRA ne met pas en avant de façon aussi accentuée les objectifs du Grenelle (traité par ailleurs par la partie recherche du Plan Ecophyto), pour autant les projets soumis sont marqués fortement par cet agenda. C'est un indicateur fort du fait que les programmes ont des effets performatifs mais que les communautés de recherche ont aussi une autonomie de positionnement scientifique de leurs objets et finalités, en lien ici avec une actualité politique qui polarise l'attention des chercheurs.

Nous avons ensuite procédé de façon systématique à l'analyse de la cooccurrence des multi-termes extraits dans les résumés des projets pour caractériser de façon fine la composition des associations de termes. Sur cette base, l'analyse des clusters obtenus permet une interprétation des regroupements significatifs d'appariements de multi-termes. Cette partition recèle des champs sémantiques marqués pour 8 clusters par des thématiques des sciences du vivant, des sciences économiques et de gestion, et pour 5 autres clusters par des thématiques des sciences sociales. Nous proposons une description de ces clusters qui indiquent des polarités des promesses scientifiques. Nous avons enfin projeté la variable « projet lauréat » sur cette partition ; le résultat indique que l'ensemble des clusters est concerné (donc pas d'effet d'éviction thématique par le financement), sauf sur des « zones » que l'on peut considérer comme des trous structuraux ( 1. Santé humaine et produits et pratique de phyto-protection; 2. Durabilité, Changement de pratiques, dynamiques spatiales et qualités des eaux/sols, 3. Sécurité alimentaire, bien public, valeurs, pays du Sud) sachant que les deux premières thématiques ont été mises en traitement dans le cadre d'autres programmes (plan Ecophyto et appel de l'INSERM).

Poursuivant notre travail de caractérisation, nous avons ensuite étudié l'évolution de la composition sémantique des résumés de ces projets dans le temps. Cela a consisté à mesurer l'évolution des nouveaux termes qui apparaissent d'une année sur l'autre, et montré que la deuxième année du programme ADD montre le maintien d'une nouveauté importante puisque plus de 100 multi-termes nouveaux apparaissent. Ce mouvement se maintient ensuite pour la première année du programme SYSTERRA. Par contre, celui-ci connaît une baisse de cette exploration, puis une reprise. On notera que le programme AGROBIOSPHERE apporte peu de nouveauté ; tout se passe comme si ce programme était plutôt porteur d'une logique d'approfondissement dans un périmètre lexical déjà fixé. Les réponses faites par les communautés de recherche semblent donc faire écho à la polarisation de l'intention exprimée dans le dernier programme. Sur cette base nous avons ensuite caractérisé la dynamique sémantique en opérant une clusterisation temporelle pour analyser les relations de dépendance temporelle des clusters d'une année à une autre. Cela a permis de noter des bifurcations ou des fusions de clusters et l'apparition de nouveaux clusters. Nous mettons ainsi en évidence des continuités et des discontinuités thématiques dans l'offre de promesses scientifiques. L'effet de la programmation est donc dual : il induit la constitution d'une offre de projets proposant des thématiques existant sur la durée du programme, mais également induit des polarisations. Nous avons caractérisé ce mouvement en restituant les thématiques permanentes et celles plus contingentes de certaines périodes.

(2) Après avoir rappelé que les projets sont des oeuvres collectives des textes scientifiques à part entière, nous avons procédé à une étude sociologique des collaborations dans ces projets. Un premier résultat concerne la mise en évidence d'un *pool* d'unités qui ont déposé fortement dans les trois programmes (10% des unités l'ont fait plus de 5 fois) et tout un ensemble d'unités nombreuses (plus de 60% du total) ne sont participantes qu'une seule fois. Cette distribution reflète une loi de puissance et n'est pas surprenante : des laboratoires de taille importantes ont plus de capacité de soumettre du fait du nombre de

chercheurs, des laboratoires peu positionnés dans le domaine peuvent rejoindre un projet du fait de collaborations particulières entre 2 chercheurs ou 2 équipes. Cette distribution indique que les incitations des programmes mettent en visibilité une structuration d'unités qui sont des attracteurs pour d'autres unités, ainsi que parfois pour des acteurs de la recherche appliquée. Enfin, trait important, on note que sur la durée des ces 3 programmes des établissements de plus en plus variés se sont trouvés impliqués à partir d'une contribution initiale à ADD très forte et structurant de l'INRA. L'étude du trend des collaborations fait état d'un mouvement d'extension qui maintient une capacité d'exploration des collaborations scientifiques. Il milite pour indiquer l'idée d'un régime de recherche en cours de stabilisation dans sa composante professionnelle. L'impact de la programmation connaît ainsi deux phases : la première auprès d'une population spécifique d'unités qui ancre nettement le travail de recherche sur la thématique du développement durable de l'agriculture, puis une phase d'expansion dans une population d'unités qui s'enrichit continuellement.

Nous avons ensuite produit des résultats plus directement sur la composition des clusters de collaborations entre les laboratoires. Nous avons repéré les unités qui sont les principaux attracteurs du programme par leur niveau de contribution et noté qu'elles ont une position souvent d'intermédiaire ou de carrefour de collaborations, même si elles contribuent de façon significative à construire un cluster de collaborations particulières. Les collaborations intra-cluster peuvent alors être interprétées comme relevant d'une logique de « club », reflétant des réseaux de collaborations significativement fortes entre des équipes en position d'équivalence structurelle dans le réseau du cluster ; les positions de certains laboratoires entre des clusters sont elles indicatrices d'un type de collaborations différentes, leur position intermédiaire traduit une position de carrefour. Partant, nous avons caractérisé la topologie de ces clusters de collaborations et indiqué l'existence d'une structuration des collaborations autour de trois clusters plus centraux portés par des laboratoires « carrefours ». Nous avons également établi que les choix de sélection (variable projet lauréat) n'avaient pas induit de façon significative une préférence pour certaines collaborations.

(3) Pour finir notre entreprise de caractérisation, nous avons procédé à l'étude des relations dites bipartites entre les laboratoires et les multi-termes. Nous avons ainsi observé l'existence d'une structuration socio-sémantique du paysage des clusters suivant quatre groupes principaux. Nous avons nommé ces groupes à partir d'une analyse de leur composition pour signifier l'impact des trois programmes dans la constitution d'un potentiel de recherche suivant quatre domaines principaux :

- Sciences de l'écologie des sols et des paysages;
- Sciences de la durabilité de l'agriculture par la régulation des activités humaines ;
- Sciences de la gestion à haute valeur environnementale des ressources ;
- Sciences de la régulation écologique des pressions biotiques et anthropiques.

En reportant les programmes sur ce paysage on met également en évidence que le programme ADD a particulièrement impacté la constitution d'une offre de Sciences de la durabilité de l'agriculture par la régulation des activités humaines; le programme SYSTERRA marque de façon significative la constitution d'une offre de Sciences de la régulation écologique des pressions biotiques et anthropiques et de Sciences de l'écologie des sols et des paysages; enfin le programme AGROBIOSPHERE impacte la constitution d'une offre des Sciences de la durabilité de l'agriculture par la régulation des activités humaines. On tire de ce travail de projection des programmes sur les agencements de laboratoire et de mots-clés, l'idée que ces programmes engendrent une logique d'offre et de portage de promesses scientifiques à la fois globale mais aussi suivant certaines intensités particulières.

Notre travail comporte certaines limites liées notamment à l'interprétation qui fait suite à la production des cartes et surtout du paramétrage des analyses (choix de la métrique, filtrage par des seuils, lecture

visuelle des graphes et sensibilité de la partition des clusters au paramétrage). Ce n'est pour autant pas un trait spécifique à notre méthodologie ni à l'outil utilisé, car de nombreuses démonstrations scientifiques reposent sur le paramétrage des instruments utilisés pour l'analyse. Une autre limite est liée à notre approche foncièrement limitée à une approche socio-sémantique de la littéracie scientifique que proposent ces projets. Le travail que nous avons réalisé devrait être ainsi complété par une description de ce que représentent, tant le travail d'écriture que les conditions et les apports des collaborations entre laboratoires et - à un niveau plus fin- entre les équipes ou même leurs principaux investigateurs. Cela requerrait une étude sociologique spécifique ainsi qu'une caractérisation des unités sur le plan de leur localisation géographique, de leur profil disciplinaire, voire de leur taille. Notre étude des traces pourrait alors être complétée par une étude sociologique des conditions et enjeux de leur production.

## Références bibliographiques

- ANR, (2015). Émergence de l'Agroécologie : perspectives pour le futur. Les programmes ADD-SYSTERRA-AGROBIOSPHERE - *Cahier de l'ANR n°8 - juillet 2015*, Paris : Agence Nationale de la Recherche.
- Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics*, 2(1), 1-52.
- Barbier M., Breucker P., Ollivier G., (2013). La Biodiversité en Textes: figures de l'institutionnalisation d'un domaine de recherche, *Rapport du WP.1 du Projet Pan-Bioptique, Programme ANR*, Marne La Vallée, Juillet 2013, 77 p
- Barbier, M., Farges, M., Batifol, V., Mogoutov, A. (2012). Textual analysis and scientometric mapping of the dynamic knowledge in and around the IFSA community. In: Ika Darnhofer, David Gibbon, Benoît Dedieu (Eds), *Farming systems research into the 21st century : The new dynamic* (p. 73-94). Life Sciences. Dordrecht, NLD : Springer Link.
- Barrier, J., (2010). *La science en projets : régimes de financements et reconfiguration du travail des chercheurs académiques. Le cas des sciences et technologies de l'information et de la communication en France (1982–2006)*. Thèse de doctorat en sociologie, Institut d'Etudes politiques de Paris.
- Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): past, present, and future, *Scientometrics*, 52(3) : 365-377.
- Börner, K., Chen, C. M., and Boyack, K.W. (2003). Visualizing knowledge domains. *Annual review of information science and technology*, 179–255.
- Bourret, P., A. Mogoutov, C. Julian-Reynier, and A. Cambrosio (2006). A new clinical collective for french cancer genetics a heterogeneous mapping analysis. *Science, Technology, & Human Values*, 31: 431–464.
- Bozeman, B., & Rogers, J. D. (2002). A churn model of scientific knowledge value: Internet researchers as a knowledge value collective. *Research Policy*, 31(5): 769-794.
- Callon, M., J. Law, and A. Rip (1986). *Mapping the dynamics of science and technology*. London: MacMillan.
- Callon, M., J.P. Courtial, W.A. Turner, and S. Bauin (1983). From translations to problematic networks: An introduction to co-word analysis, *Social Science Information*, 22: 191–235.
- Cambrosio A., Keating P., Mogoutov A. (2004). Mapping collaborative work and innovation in biomedicine: A computer assisted analysis of antibody reagent workshops. *Social Studies of Science*, 34: 325–364.
- Cambrosio, A., Cotterau P., Popowycz S., Mogoutov A., and Vichnevskaja T. (2010). Analysis of heterogenous networks: The RéseauLu Project. In: C. Brossaud and B. Reber (Eds.). *Digital cognitive technologies. Epistemology and knowledge society*, London: ISTE/Wiley, pp. 137–152.

- Cambrosio, A., Keating P., Mercier S., Lewison G., and Mogoutov A. (2006). Mapping the emergence and development of translational cancer research, *European journal of cancer*, 24: 3140-3148
- Cetro R., Barbier M., Breucker P., Eggermont H., Gambette P., Kyriacopoulou T., Le Roux X. Martineau C., Turenne N. (2014). Vers une approche semi-automatique pour la définition de motifs d'argumentation utilisés dans les résumés de projets scientifiques du domaine de la biodiversité, *RNTI-SHS-2*, pp.47-80.
- Chateauraynaud F., Reber B. et Van Meter K., (2003). Marlowe, Prospero et la technologie littéraire, *Bulletin de Methodologie Sociologique*, 79 : 5-46.
- Chavalarias D., Cointet JP. (2008). Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study, *Scientometrics*. 75(1): 37-50.
- Chavalarias D., Cointet J-P., (2013). Phylomemetic Patterns in Science Evolution : The Rise and Fall of Scientific Fields, *PloS one*, 8(2) : 54847
- Cointet J-P. (2009). *Dynamiques sociales et sémantiques dans les communautés de savoirs: Morphogenèse et diffusion*, Thèse de l'Ecole Polytechnique.
- de Saint Léger M. , et van Meter K.M. (2005) Cartographie du premier congrès de l'ASF avec la méthode des mots associés, *Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique*, 85.
- Ding, Y., Gobinda G. Chowdhury, Schubert Foo, 2001. Bibliometric cartography of information retrieval research by using co-word analysis, *Information Processing and Management*, 37 : 817-842.
- Gingras, Y., 2008. La Fièvre de l'évaluation de la recherche du mauvais usage de faux indicateurs, *Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique*, 100 (1) : 41-44.
- Gusfield J. (1976). « The literary rhetoric of science: comedy and pathos in drinking driver research », *American sociological review*, 41: 16-34.
- He, Q. (1999). *Knowledge Discovery through Co-Word Analysis*. Library Trends.
- Hoekman, J., Scherngell. T., Frenken, K. and Tijssen, R., 2013. Acquisition of European research funds and its effect on international scientific collaboration, *Journal of Economic Geography*, 13 : 23-52
- Hubert M., Chateauraynaud F., Fourniau, J.-M., (2012). Les chercheurs et la programmation de la recherche : du discours stratégique à la construction de sens, *Quaderni*, 77 (1) : 85-96.
- Hubert M., Louvel S., (2012). Le financement sur projet : quelles conséquences sur le travail des chercheurs ?, *Mouvements*, 71 (3) : 13-24.
- Jenny, J., (1997). Méthodes et pratiques formalisées d'analyse de contenu et de discours dans la recherche sociologique française contemporaine. Etat des lieux et essai de classification, *Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique*, 54 (1) : 64-122
- Jones, D.S., Cambrosio, A., and A. Mogoutov A., (2011). Detection and characterization of translational research in cancer and cardiovascular medicine. *J Transl Med*. 9: 57-69
- Jouvenet M., (2011). Profession scientifique et instruments politiques : l'impact du financement "sur projet" dans des laboratoires de nanosciences, *Sociologie du travail*, 53(2) : 234-252.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration?. *Research policy*, 26(1), 1-18.
- Landry J., Traore N., Godin B., (1996). An econometric analysis of the effect of collaboration on academic research productivity, *Higher Education*, 32: 283-301.
- Larédo P., (2010). Le financement de la recherche sur projets en France : positionner pour mieux questionner, in Hoillard M., Saez G. (dir), *Politique, science et action publique*, PUG, p. 177-196.
- Larédo P., Mustar P. (Eds.). (2001). *Research and innovation policies in the new global economy : An international comparative analysis*, Edward Elgar Pub.
- Latour B. et Fabbri P. (1977). Pouvoir et devoir dans un article de science exacte, *Actes de la Recherche en Sciences Sociales*, (13): 81-95.

- Laudel, G., (2006). The Art of Getting Funded: How Scientists Adapt to their Funding Conditions. *Science and Public Policy*, 33 (7), 489–504.
- Law, J., Bauin, S., Courtial, J., & Whittaker, J. (1988). Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification, *Scientometrics*, 14(3-4), 251-264.
- Lazega E., Mounier L., Stofer R., Tripier A., (2004). Discipline scientifique et discipline sociale : Réseaux de conseil, apprentissage collectif et innovation dans la recherche française sur le cancer (1997-1999), *Recherches sociologiques*, 35(3) : 3-27.
- Lee S. and Bozeman B., (2005), The Impact of Research Collaboration on Scientific Productivity, *Social Studies of Science*, 35 (5) : 673-702.
- Lepori, B., van den Besselaar, P., Dinges, M., van der Meulen, B., Poti, B., Reale, E., Slipersæter, S., Thèves, J., 2007. Indicators for comparative analysis of public project funding: concepts, implementation and evaluation, *Research Evaluation* 16 (4), 243–255.
- Leydesdorff, L. and I. Hellsten (2006). Measuring the meaning of words in contexts: An auto-mated analysis of controversies about 'Monarch butterflies,' 'Frankenfoods,' and 'stem cells', *Scientometrics* 67: 231–258.
- Louvel S., (2007). « Le nerf de la guerre », *Revue d'anthropologie des connaissances*, 1 (2), p. 297-322.
- Lucio-Arias, D. and L. Leydesdorff (2007). Knowledge emergence in scientific communication: From "fullerenes" to "nanotubes", *Scientometrics* 70: 603–632.
- Miller J.D., 1983. Scientific Literacy: A Conceptual and Empirical Review, *Daedalus*, 112 (2) : 29-48
- Mogoutov, A., & Kahane, B. (2007). Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking, *Research Policy*, 36(6), 893-903.
- Noyons, E. (2001). Bibliometric mapping of science in a policy context, *Scientometrics* 50: 83–98.
- O'Neill J. (1981). The literary production of natural and social inquiry : issues and applications in the social organization of science, *Canadian Journal of Sociology*, 6, (2).
- Paradeise, C., and Thoenig J-C., (2013). Academic institutions in search of quality. Local orders and global standards, *Organization studies*, 34(2), 195-224.
- Pontille D., and Torny D., (2010). The controversial policies of journal ratings: evaluating social sciences and humanities, *Research Evaluation*, 19 (5): 347-360.
- Powell, W.W., D.R. White, K.W. Koput, and J. Owen-Smith (2005). Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. *American Journal of Sociology* 110: 901–975.
- Price, D. J. de S. (1976). A General Theory of Bibliometric and Other Cumulative Advantage Processes. *Journal of the American Society for Information Science and Technology*, 27(5-6), 292–306.
- Raimbault, B., Cointet, J. P., & Joly, P. B. (2014). Caractérisation du processus d'émergence de la biologie synthétique à partir d'une approche scientométrique. *Médecine/Sciences*, 29, 47-55.
- Rip A. (1988). Mapping of science: possibilities and limitations, in A.F.J. van Raan (ed), *Handbook of Quantitative Studies of Science and Technology*, Elsevier Science Publishers B.V., 253-27.
- Roth, C. and J.-P. Cointet (2010). Social and semantic coevolution in knowledge networks. *Social Networks* 32: 16–29.
- Schultz E., (2013). Le temps d'un projet, *Temporalités*, 18, mis en ligne le 19 décembre 2013, consulté le 15 juin 2015. URL : <http://temporalites.revues.org/2563>.
- Sewell, W., H., (1992). A Theory of Structure: Duality, Agency, and Transformation, *American Journal of Sociology*, 98(1): 1-29
- Small, H. (1973). Co-citation in scientific literature: A new measure of the relationship between publications. *Journal of the American Society for Information Science*, 24: 265–269.

Tancoigne E., Barbier M., Cointet J-P., and Richard G., (2014). The place of agricultural sciences in the literature on ecosystem services, *Ecosystem Services*, 10 : 35-48.

Tari, T. (2015). Savoirs contés. *Réseaux*, 188(6) : 53-83.

Tari, T., P. Breucker, and M. Barbier (2010). Characterising the localisation of projects collaborations in research dynamics: Methodological requirements and results for new visualisations of heterogeneous networks. *Paper presented at the ENID conference*.

Thèves J., Lepori B., Larédo P., (2007). Changing patterns of public research funding in France, *Science and Public Policy*, 34 (6) : 389-399.

van Raan, A. (2000), On growthn ageing, and fractal differentiation of science, *Scientometrics*, 47 : 347–362.

## Annexe 1. Présentation de la plateforme CorText



### Une plateforme pour l'analyse de corpus textuels :

Assurer le développement de recherches portant sur les sciences, les techniques et l'innovation en société suppose de relever des défis de construction, de traitement et d'analyse de jeux de données numériques dont la prolifération est croissante et la variété considérable. Comme l'indique la notion d'Humanité Numérique, il est devenu important aujourd'hui de mettre à profit l'étude d'un nombre croissant de bases de données numériques permettant une exploration à grande échelle de la littérature scientifique, de la littérature grise, de la production et de la gestion des données des laboratoires, des brevets, des articles de presse, de la blogosphère,...

Mais cette profusion de données n'a d'intérêt que si sont mises au point de nouvelles méthodes d'exploration et d'analyse en lien avec les questions que se posent les chercheurs de différentes disciplines des sciences humaines et sociales. Le projet CorText est une réponse à ce défi.

### Objectif

L'objectif de la plateforme CorText-Manager (<http://manager.cortext.net>) est de mettre à disposition pour le travail de recherche en SHS, des outils, des chaînes de traitement, des procédures et des méthodes pour aider les chercheurs à traiter, caractériser, analyser et quantifier des données textuelles organisées, plus ou moins calibrées. Pour cela, l'équipe CorText met à disposition des compétences, des outils et une interface en ligne pour deux types d'approches : L'analyse numérique de données et les analyses distributionnelles et relationnelles.

### Une spécialité : l'étude et la caractérisation de domaine de recherche

Tracer et cartographier la structure, l'organisation interne et la dynamique de production de connaissances grâce à la disponibilité de grandes bases de données représente un enjeu important pour plusieurs disciplines qui héritent de l'Intelligence Artificielle et la statistique textuelle.

### L'Equipe CorText-Lab :

Marc BARBIER, Directeur de recherche, Directeur de la Plateforme, INRA

Bilel BENBOUZID, maître de Conférence, UPEM

Philippe BREUCKER, Ingénieur, INRA

Jean-Philippe COINTET, Ingénieur de recherche (IPEF), INRA

Constance DE QUATREBARBES, Ingénieur, UPEM

Patricia LAURENS, chargée de recherche, CNRS

Antoine MAZIERES, CJS, INRA

Guillaume ORSAL, Ingénieur, INRA

Nicolas TURENNE, Chargé de recherche, INRA

Antoine SCHOEN, enseignant chercheur, ESIEE

Elise TANCOIGNE, Post-doctorante, UPEM

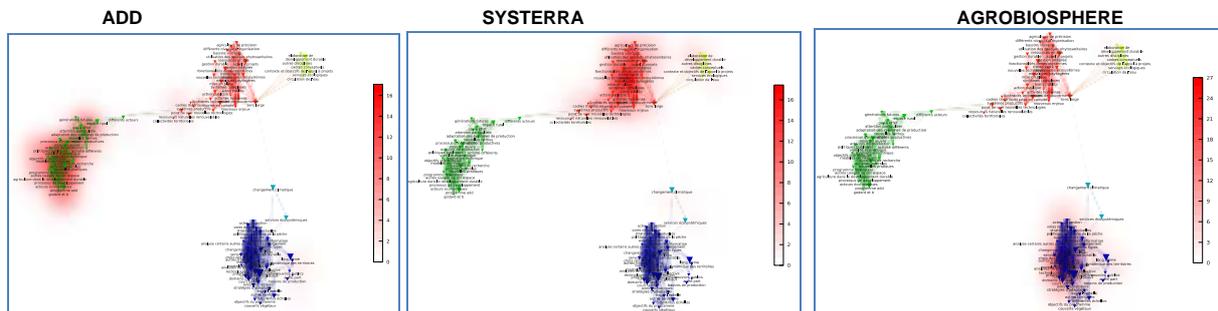
Lionel VILLARD, enseignant chercheur, ESIEE

*Avec les contributions de Andréi MOGOUTOV (EXVISU), Tam Kien DUONG (FORMISM) et Cloé DULOQUIN (Graphiste),*

### Avec le soutien de :

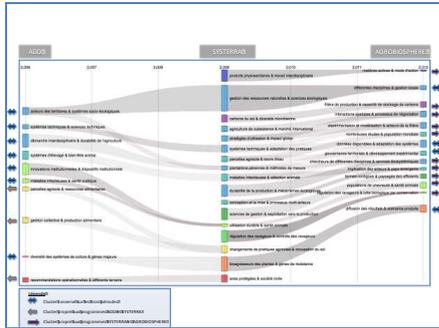


## Annexe 2 : Clusterisation et tableaux des N-Gram extraits des textes des Appels à Projets (multi-termes spécifiques des appels)

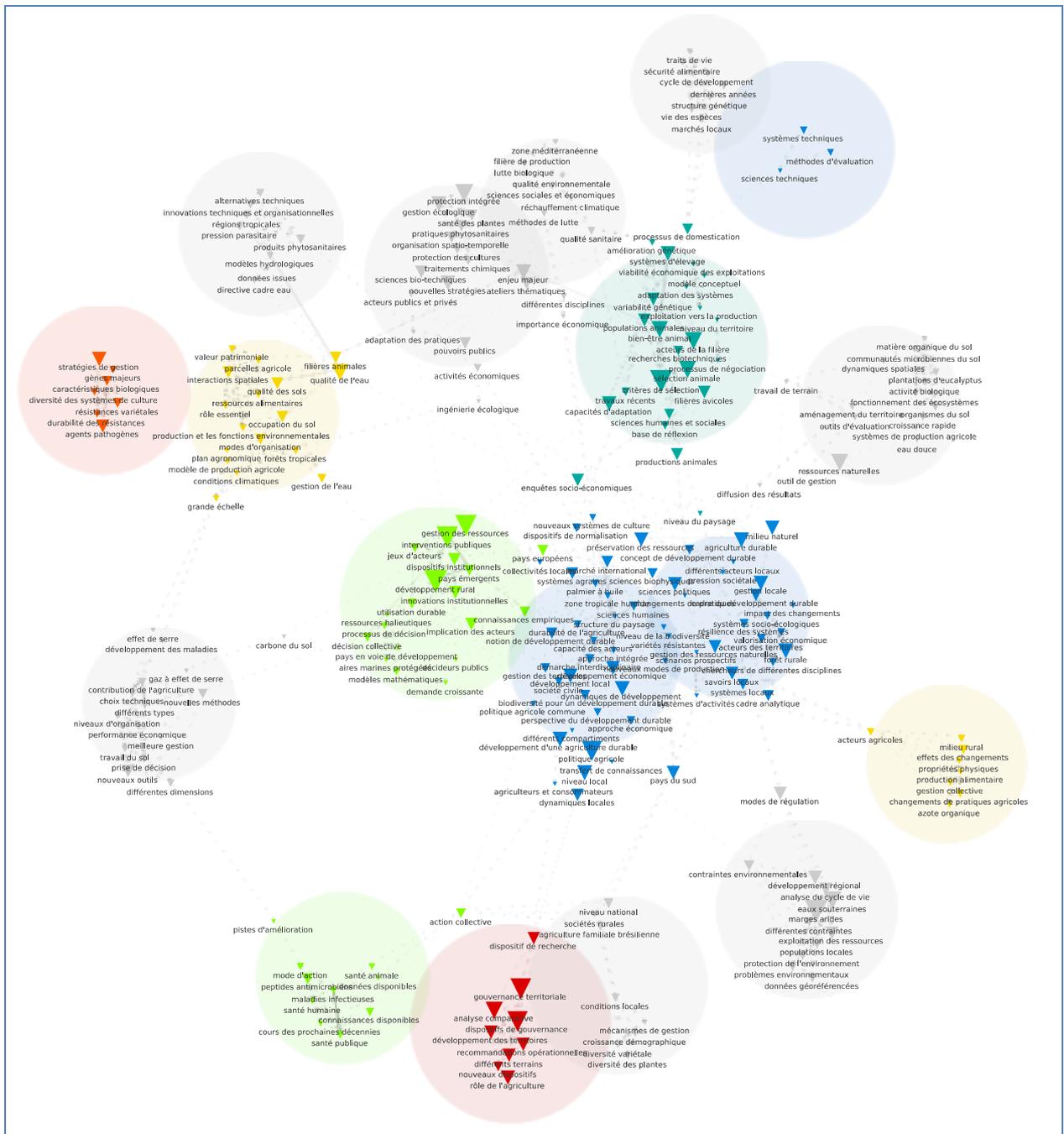


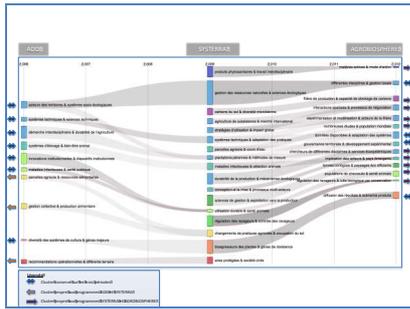
ADD	SYSTERRA	AGROBIOSPHERE
objectifs scientifiques	prise en compte	biens et services
modèles techniques	action publique	systèmes d'information
Godard et B	activités humaines	axes thématiques
innovation technique	itinéraires techniques	différents types
activités de production	nouvelles techniques	pouvoirs publics
acteurs économiques	modes de gouvernance	dynamique des territoires
nouvelles pratiques	utilisation des produits phytosanitaires	voies de solution
espace rural	systèmes complexes	différentes échelles
processus de développement	nouveaux enjeux	analyse certains autres processus de changement
autres usages de cet espace	systèmes productifs	action collective
agriculture dans le développement durable	agriculture de précision	bassins de production
différents acteurs	ressources en eau	choix des cultures
processus d'innovation	fonctionnalités des écosystèmes	autre part
viabilité économique	gestion durable	eau à l'échelle
nouvelles formes	sens large	mortalité des abeilles
entités productives	mise en place	couverts végétaux
programme fédérateur	biens publics	autres usages
activités agricoles	services des écosystèmes	technologies de l'information
gaz à effet	bassins versants	actes de gestion
activité différents	collectivités territoriales	analyses de trajectoires
générations futures	ressources naturelles renouvelables	court terme
questions de recherche	nouvelles technologies	autres transformations
mise en œuvre	trame verte	cycle de l'eau
attention particulière	point de vue	autres territoires
politiques publiques	différents niveaux d'organisation	long terme
adaptation des systèmes de production	appel à projets	caractère durable
	gestion intégrée	changement et tendances
	mosaïques paysagères	service d'une gestion durable
	cadres théoriques	politique commune de la pêche
		stratégies individuelles
		stratégies d'adaptation

### Annexe 3 : clusterisation des termes des résumés pour les 3 programmes

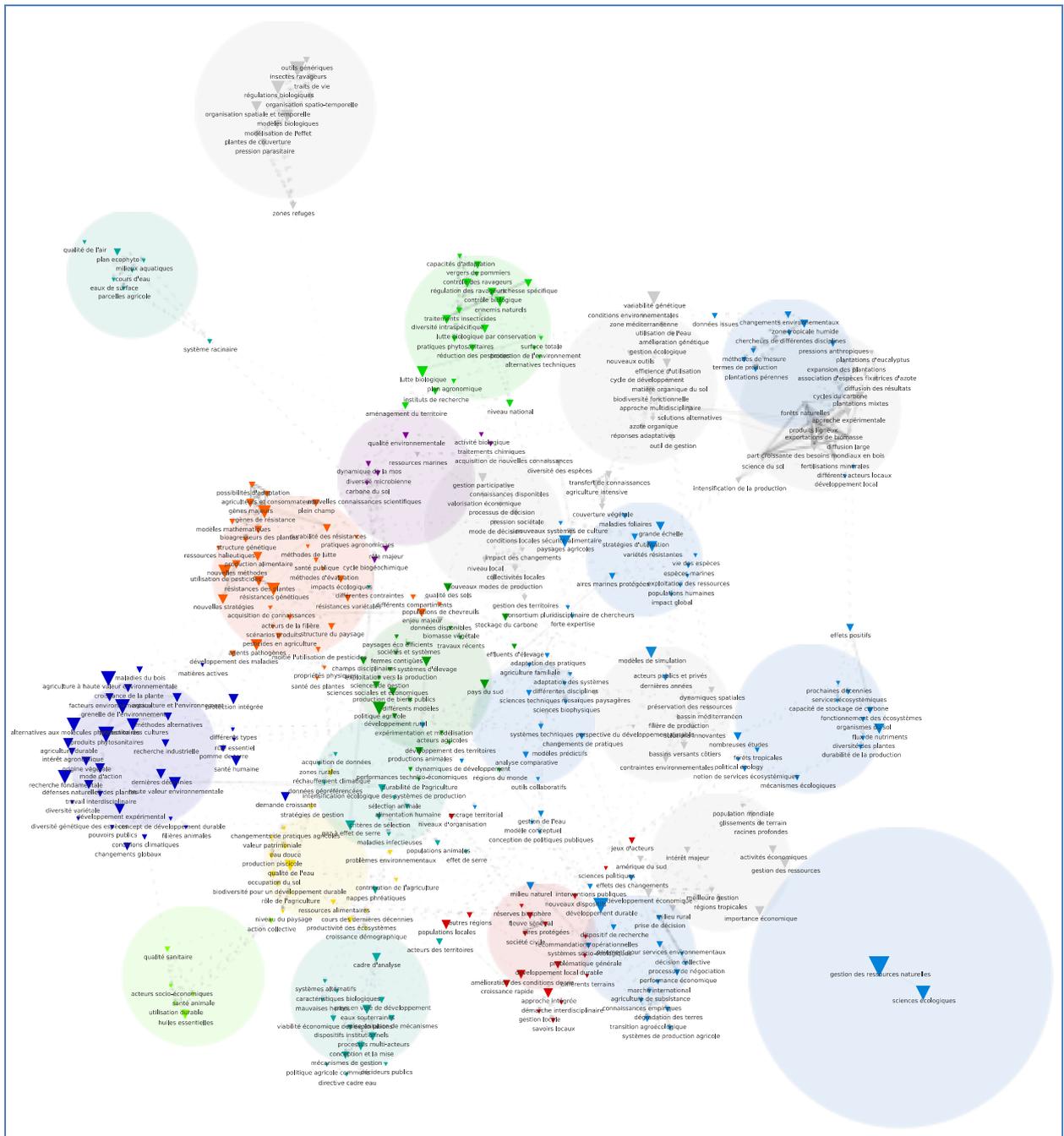


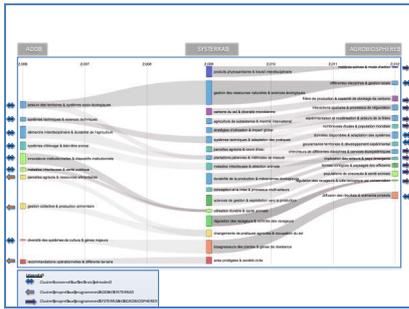
### Programme ADD



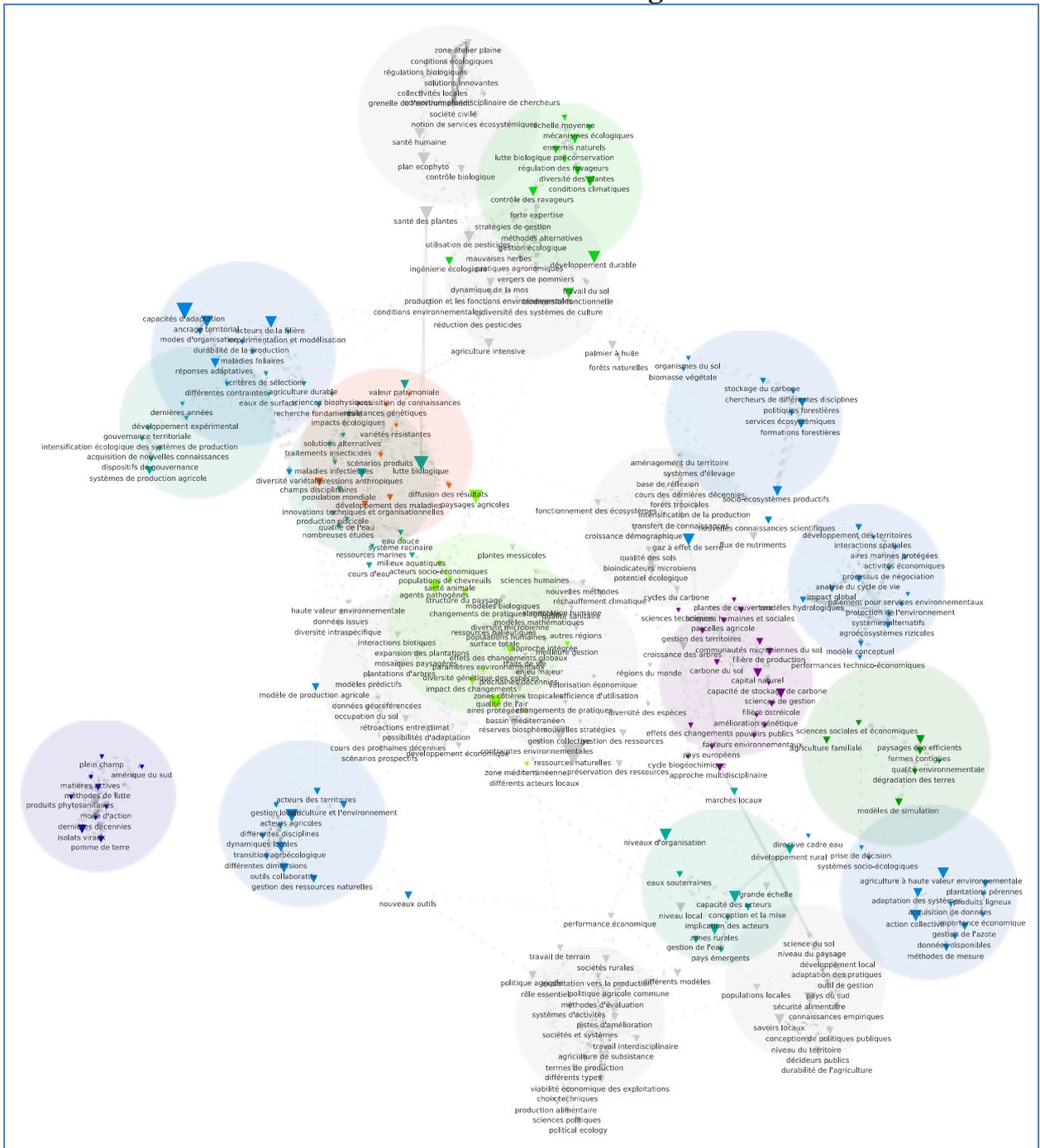


# Programme SYSTEMERRA





# Programme AGROBIOSPHERE







AGENCE NATIONALE DE LA RECHERCHE  
**ANR**  
[www.anr.fr](http://www.anr.fr)

 **INRA**  
[www.inra.fr](http://www.inra.fr)