

# (Comment) peut on isoler la subjectivité de l'auteur dans une analyse des données textuelles? Etude de la variabilité du vocabulaire dans les atlas français des paysages

Congrès AFS 2013, RT Méthodes, 05.09.2013

*Julie Bourbeillon, UMR IRHS*

*Jonathan Charles, UMR GRANEM*

*Damien Rousselière, UMR GRANEM*

*Julien Salanié, UMR GATE*

*Djery Sow, UMR GRANEM*



**PAYTAL**  
**PAYSAGE ET**  
**ETALEMENT URBAIN**



# Contexte de la présentation : le projet PAYTAL

---

*Introduction*

Méthodologie

Présentation  
des données

Quelques  
résultats

Prolongements

- PAYTAL : Paysage et Etalement Urbain – Dynamiques physique et culturelle et modélisation
- Programme financé par le MEDDE (2012-2014) « Paysage et Développement Durable » PDD2
- Equipe pluridisciplinaire : Economie, Sociologie, Géographie Sociale, Géographie physique, Informatique, Bioclimatologie et Télédétection
- Etude de la dynamique conjointe de l'étalement urbain et des paysages
  - Rôle des paysages et de leur modification sur l'étalement urbain
  - Rôle de la dynamique urbaine dans la modification des paysages

# Contexte de la présentation : le projet PAYTAL

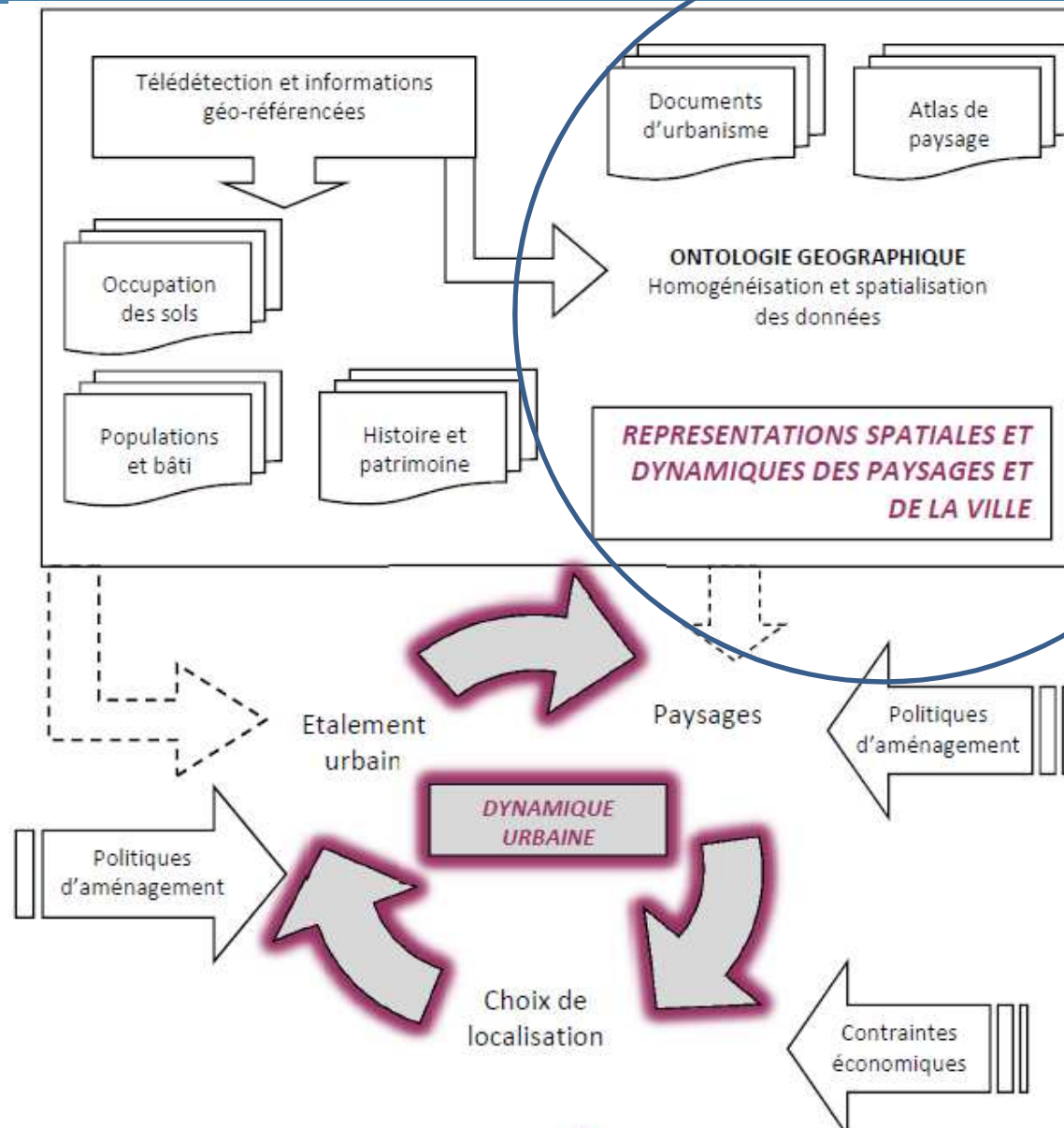
Introduction

Méthodologie

Présentation  
des données

Quelques  
résultats

Prolongements



# Objectif : qualifier la perception des paysages

Introduction

Méthodologie

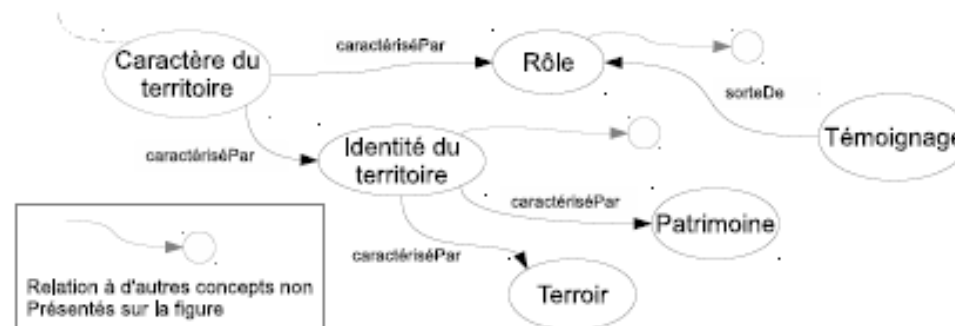
Présentation  
des données

Quelques  
résultats

Prolongements

Convention Européenne du Paysage (2000): « le paysage est « une partie de territoire telle que perçue par les populations, dont le caractère résulte de l'action de facteurs naturels et/ou humains et de leurs interrelations »

⇒ Idée de constitution d'une ontologie du paysage (e.g. Cataldo & Rinaldi 2010 ( Computers, Environnement & Urban System))



*Extrait ontologie PAYTAL*

# Objectif : qualifier la perception des paysages

Introduction

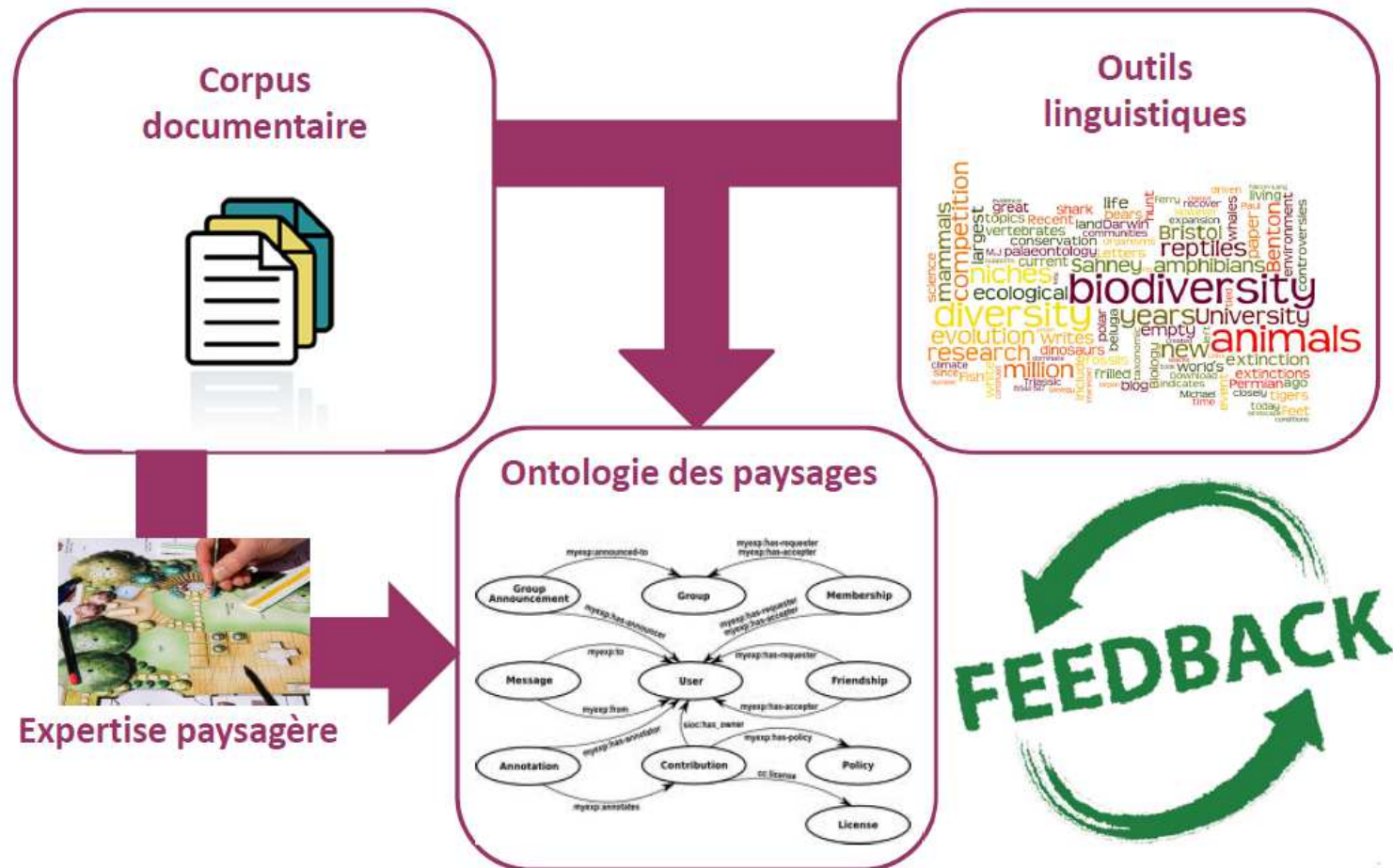
Méthodologie

Présentation  
des données

Quelques  
résultats

Prolongements

## Méthode constitution ontologie (inspiré de Bourigault & Aussenac-Gilles 2003)



# Objectif : qualifier la perception des paysages

---

*Introduction*

Méthodologie

Présentation  
des données

Quelques  
résultats

Prolongements

Corpus documentaire:

- Documents d'urbanisme et de planification (PADD - SCoT, PLU)
- Documents judiciaires (conflits judiciarisés, etc.)
- Documents de gestion (Chartes de PNR, etc.)
- Atlas de paysages « document dressant l'état des lieux des paysages et dynamiques qui les transforment » (Luginbühl 1994)

⇒ Choix

- ⇒ Sur un même lieu (agglomération angevine), traitement de l'ensemble des documents
- ⇒ Sur l'ensemble de la France : traitement des atlas de Paysages

# L'Atlas de paysages

Introduction

48 en France (entre 1994 et 2013) Environ 1500 unités paysagères « homogènes »

Méthodologie

Ex. Loire Atlantique /<http://www.paysages.loire-atlantique.gouv.fr>

Présentation des données

Atlas des Paysages de la Loire-Atlantique - Mozilla Firefox

Atlas des Paysages de la Loire-Atlantique

[www.paysages.loire-atlantique.gouv.fr](http://www.paysages.loire-atlantique.gouv.fr)

Atlas des paysages de Loire-Atlantique

Présent pour l'avenir

Rechercher OK

Accueil

- Préambule
- Présentation
- Partenaires
- Les Paysages de Plateaux
- Les Paysages Ligériens
- Les Paysages Urbains
- Les Paysages Littoraux et de Marais
- Observatoire des paysages
- Ressources
- Carte de synthèse des Unités Paysagères

Les Unités Paysagères par commune

Choix de la commune

PAYSAGES de PLATEAUX

PAYSAGES LIGERIENS

PAYSAGES URBAINS

Quelques résultats

Prolongements



## Les marais de Brière

«Un soleil ardent tombait sur la Brière. De Fédrun aux coteaux des rives, ce n'était qu'un grand lac de lumière torrentielle. De tous côtés sous de tremblantes vapeurs, scintillait le feu des coulins. Et pas un pouce d'ombre au large ; pas un bruit non plus, que de temps à autre, du fond des lointains, comme un vague écho, la chanson des broyeuses de chanvre.» Alphonse de Chateaubriand 1923, la Brière

Les limites paysagères de cette sous-unité sont principalement constituées par le coteau bocager du sillon de Bretagne, l'interface boisée avec le plateau de Guérande à l'ouest et au sud la limite est beaucoup plus anthropique (voie ferrée, urbanisation et route)



C'est une topographie qui joue finement sur l'horizontalité et la présence de l'eau qui conditionne la structuration des paysages: les terres basses inondables, les îles, les plateaux avec des franges de coteau souvent marquées. Sur cet espace plan, la combinaison des configurations naturelles ou agricoles variées composent des paysages à la végétation contrastée.

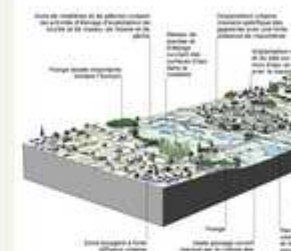


✉ Envoyer par courriel

🖨 Imprimer

📄 Exporter au format PDF

### Bloc-diagramme



Les marais de Brière

### Carte de la sous-unité paysagère



Les marais de Brière

### Carte de localisation de la sous-unité dans l'unité





# L'Atlas de paysage : un objet problématique

---

*Introduction*

Méthodologie

Présentation  
des données

Quelques  
résultats

Prolongements

Méthodologie de constitution (Brunet-Vinck 2004) :

« la caractérisation des paysages ne doit pas se contenter d'une description de l'utilisation du sol, mais tenter de restituer les caractères de l'ensemble de l'aspect du territoire et de ce que l'on en perçoit »

⇒ Idée d'une analyse de la dimension « sensible » du paysage (Davodeau 2009)

⇒ subjectivité de l'auteur mise en avant / droit à l'interprétation

⇒ « résolue » par la constitution d'une équipe pluridisciplinaire (multiplication des points de vue)

# Méthodologie

---

Introduction

*Méthodologie*

Présentation  
des données

Quelques  
résultats

Prolongements

Deux travaux en parallèle:

- Construction semi-automatique (experte et automatique) d'une construction d'une ontologie (IC 2013) (notamment car importance des événements rares)
- Tenter de quantifier/appréhender l'importance de la subjectivité

# Quantifier l'importance de la subjectivité

Introduction

Méthodologie

Présentation  
des données

Quelques  
résultats

Prolongements

Trois possibilités

⇒ Analyse multiniveau (à classification croisée non hiérarchique) (Yang & Land 2006, 2008)

⇒ Exploiter le fait que les auteurs réalisent plusieurs atlas (seul ou à plusieurs)

⇒ Ajouter des données externes (e.g. Corine Land Cover) comme variables de contrôle



⇒ Pas assez de variabilité (26/48 atlas traités)

# Quantifier l'importance de la subjectivité

---

Introduction

*Méthodologie*

Présentation  
des données

Quelques  
résultats

Prolongements

Trois possibilités

⇒ Appariement (données expérimentales)

⇒ Exploiter le fait que les unités paysagères  
« débordent » les atlas et peuvent être  
coupées par des limites administratives  
(...qu'on espère exogènes)

⇒ Deux atlas (avec auteurs ou ensembles  
d'auteurs différents) étudient la même zone

⇒ .. Éventuellement étendue aux plus proches  
voisins (économétrie spatiale)

⇒ En cours...

⇒ Problème de géospatialisation des  
données

# Quantifier l'importance de la subjectivité

---

Introduction

*Méthodologie*

Présentation  
des données

Quelques  
résultats

Prolongements

Trois possibilités

⇒ Données quasi-expérimentales

⇒ Identification des compétences des auteurs

⇒ Hypothèse : choix des compétences  
indépendant des caractéristiques  
*inobservables* de la zone à décrire

⇒ Identification de la variabilité des thématiques  
(dictionnaires *ad hoc*, scénario « concepts » de  
Tropes )

⇒ Identification de la variabilité des termes au  
sein de chaque thématique (fréquences,  
indices d'entropie (shannon))

# Présentation des données

Introduction

Méthodologie

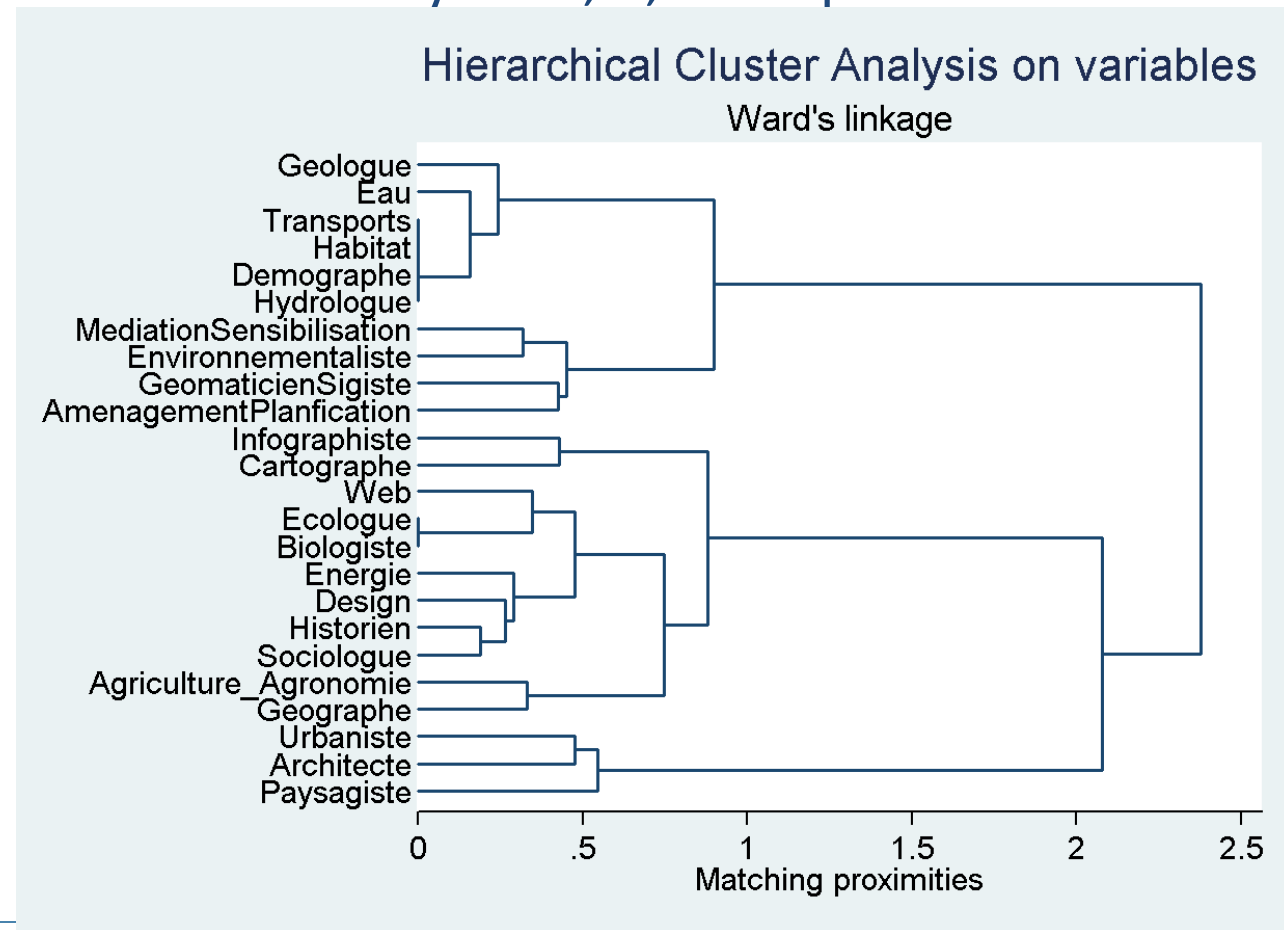
*Présentation  
des données*

Quelques  
résultats

Prolongements

Numérisation réussie de 26 Atlas soit 625 unités  
paysagères

1,5 auteurs en moyenne, 5,4 compétences



# Résultats

---

Introduction

Méthodologie

Présentation  
des données

*Quelques  
résultats*

Prolongements

⇒ **Identification de la variabilité des thématiques (dictionnaires ad hoc, scénarios Tropes)**

⇒ Analyse de données (ACM + CAH)

⇒ Régression sur fractions (fractional multinomial logit) (Papke & Wooldridge 1996, Buis 2008...)

⇒ Régression Dirichlet / sur Simplex.. (Aitchison)

⇒ **Identification de la variabilité des termes au sein de chaque thématique (fréquences, indices d'entropie (shannon))**

⇒ Analyse de données (ACM + CAH)

⇒ Système de régressions linéaires simultanées (Seemingly Unrelated Regression System) (Zellner 1962)

# Résultats

---

Introduction

Méthodologie

Présentation  
des données

*Quelques  
résultats*

Prolongements

## ⇒ **Fractional multinomial logit**

- ⇒ Extension du logit multinomial aux fractions (méthode QMLE)
- ⇒ Modélisation de la répartition du vocabulaire en thématiques
- ⇒ Variables de compétences prédisent bien la répartition
- ⇒ différences importantes suivant la composition du groupe d'auteurs (calcul des effets marginaux)
- ⇒ Effet du changement de méthodologie ?
  - ⇒ Période (avant/après 2004) : +2% agriculture, -4% Autres concepts, +3% géographie



# Résultats : fractional multinomial logit

	Agriculture et environnement	Animaux et végétaux	Arts et culture	Caractéristiques	Communication et médias	Comportements et sentiments	Forces et quantités	Politique et société	Transport saéronautique et espace	Autres concepts
annee2004	-0.0453	0.247***	-0.0124	-0.181	-0.663***	-0.0795	-0.386***	-0.288	-0.502	-0.332**
totalcompetences	0.403**	0.259**	-0.250	0.261	-0.151	0.273	-0.0584	-0.572**	0.296	-0.0571
totalauteur	-0.160***	-0.00300	-0.0434	-0.110**	-0.108**	-0.135*	-0.0152	-0.229***	-0.228***	-0.165***
Amenagement	0.943**	0.296*	-0.584	0.929**	2.331***	1.790***	0.396	1.576**	1.357**	1.145***
Architecte	-0.324	-1.545***	-0.571	-0.483*	0.542*	-0.704**	0.359	1.121***	1.047*	0.00208
Urbaniste	-0.675*	0.762***	0.662*	-0.239	-0.982**	0.0967	0.0207	0.429	-0.782	-0.118
Geographe	0.213	-0.279**	0.122	0.107	1.320***	0.0715	0.528**	1.508***	1.623***	0.441**
Historien	1.739***	-0.352**	-0.349	1.532***	3.037***	1.330**	1.028**	2.342***	3.150***	1.403***
Environnementaliste	-2.005***	-0.442*	0.840	-1.418***	-1.235**	-1.849***	-0.564	-0.917	-3.321***	-0.730*
Biologiste	-1.250*	-0.989**	0.573	-0.702	-1.203	-0.510	0.216	2.335**	0.555	0.400
Cartographe	-0.417	-0.734***	0.227	-0.399	-0.0540	-0.884***	0.0769	1.107***	0.375	0.222
Geomaticien SIGiste	0.713**	0.0251	-0.00603	0.721**	1.617***	0.928**	0.597*	1.323**	0.820	0.968***
Infographiste	-0.0912	0.219***	0.574***	0.0397	1.104***	-0.116	0.111	0.792***	-0.532**	0.322***
Hydrologue	-3.465***	-3.490***	0.768	-2.753**	-0.169	-3.719**	0.460	3.807**	-0.896	-0.765
Agriculture Agronomie	0.837**	0.112	-0.108	0.915***	0.976**	1.471***	0.457	2.018***	1.589***	1.068***
Design	-1.143***	-1.788***	0.157	-0.536**	1.509***	-0.0187	0.509**	0.402	-0.751	0.0172
Energie	-0.637*	-1.186***	0.291	-0.605	-0.0187	-0.975*	-0.457	0.997**	-0.0951	-0.289
Web	-1.962***	-1.065***	0.681**	-1.625***	-0.633	-1.980***	-0.304	-0.723	-2.384***	-1.209***
Constante	-0.248	-1.675***	-1.095***	-0.0484	-1.767***	-1.573***	-1.610***	-0.572**	-2.617***	0.898***

# Résultats

Introduction

Méthodologie

Présentation  
des données

*Quelques  
résultats*

Prolongements

⇒ **SUREG**

⇒ Modèle GLS

⇒ Indépendance des équations rejetée  
(chi2(66)=1404)

⇒ Corrélation des résidus (existence de variables  
inobservées influençant la diversité du  
vocabulaire)

	Archit- ecture	Bota- nique	Econo- mie	Elevage	Minera- logie	Urba- nisme	Sylvie- culture	Geo- logie	Cam- pagne	Viti- culture	Reli- gion	eau
architecture	1,000	0,261	0,265	0,283	0,246	0,057	0,095	0,161	0,227	0,247	0,446	0,162
botanique		1,000	0,076	0,286	0,123	-0,170	0,416	0,122	0,229	0,113	0,313	0,094
economie			1,000	0,108	0,120	0,201	-0,008	-0,005	0,057	0,113	0,278	0,165
elevage				1,000	0,145	-0,034	0,173	0,171	0,427	0,154	0,288	0,124
minéralogie					1,000	0,118	-0,020	0,218	0,121	0,196	0,224	0,074
urbanisme						1,000	-0,097	-0,115	-0,003	-0,088	-0,096	0,283
sylviculture							1,000	0,117	0,148	-0,031	0,096	0,169
géologie								1,000	0,092	0,053	0,104	0,055
campagne									1,000	0,122	0,197	0,120
viticulture										1,000	0,215	-0,053
religion											1,000	0,145
eau												1,000

# Résultats : SUREG

	Archite- cture	botanique	economie	elevage	Minera- logie	Urba- nisme	Sylvie- culture	geologie	campagne	viticulture	religion	eau
Paysagiste	-0.260	0.867***	0.740***	0.0814	0.437**	-0.867***	0.489**	0.819***	0.0905	0.820***	0.303	0.0258
Amenagement Planfication	5.576***	0.893	4.128***	-1.390	0.754	0.240	-1.515	2.804**	2.769**	0.877	2.069	-0.271
Architecte	2.025***	-0.848	1.247**	-0.113	-0.117	0.555	-1.336**	0.635	1.863***	0.288	0.297	-0.199***
Urbaniste	-2.021***	0.619	-2.003***	0.281	0.567	0.00496	2.037***	-0.605	-1.966***	-0.389	0.169	0.103
Geographe	1.892***	-0.706*	0.890**	0.0332	0.200	0.183	-0.664	0.605	1.521***	0.181	0.548	-0.00764
Historien	8.392***	1.253	4.132**	0.491	1.960	0.307	-2.161	4.577**	6.950***	3.399*	2.421	-0.358
Environnementaliste	-7.976***	-1.032	-3.922**	-0.606	-2.703	-0.0919	0.849	-5.203***	-7.113***	-3.531*	-3.367*	0.452*
Biologiste	5.377***	0.135	4.028***	-0.236	3.558***	-1.798	0.413	2.679**	5.278***	4.362***	2.453**	-0.806***
Hydrologue	11.91***	-0.882	9.269***	3.548	7.202**	-3.739	-1.758	5.946**	15.30***	12.71***	6.348**	-1.747***
Geologue	-12.36***	-2.114	-9.200***	-2.581	-6.419**	2.579	0.155	-7.012***	-12.33***	-9.571***	-5.931**	1.292***
Cartographe	-1.955***	-0.890**	-1.258***	-0.954***	-1.799***	0.200	-1.144***	-1.904***	-1.943***	-1.030**	-1.414***	0.156***
Geomaticien Sigiste	7.306***	2.429*	4.583***	1.617	2.335	0.0147	0.209	4.949***	6.101***	3.159**	3.031**	-0.262
Infographiste	0.801***	0.621***	0.675***	0.424**	0.497*	0.504*	-0.303	0.394*	0.935***	0.360	0.560**	0.0691**
Agriculture Agronomie	9.365***	1.733	5.506***	0.284	2.930	-0.526	-1.282	5.754***	7.343***	5.305**	4.360**	-0.755***
Design	-0.434	-1.598***	-0.387	-0.995***	-1.389***	0.506	-1.706***	-0.221	-1.188***	-0.566	-0.947***	0.0231
Energie	8.087***	-0.462	3.545*	1.679	4.484**	-3.262*	-0.0616	4.510***	7.938***	7.947***	3.600**	-1.086***
Mediation Sensibilisation	-3.367***	-0.170	-2.298***	-1.390**	-1.403*	1.368*	-0.150	-1.121*	-2.985***	-3.157***	-2.375***	0.378***
Web	-1.728***	-0.938	-1.625**	2.193***	0.399	-0.759	1.179*	-1.349**	-0.449	1.029	0.126	0.0111
totalauteur	-0.963***	-0.226	-0.742***	0.0745	-0.228	0.126	0.172	-0.405*	-0.651***	-0.517**	-0.346	0.0864***
Annee	-0.111***	0.0222	0.0140	-0.0739***	0.0335	0.0148	-0.00142	-0.0450*	-0.0530**	-0.0231	-0.00878	-0.00809**
Constant	223.6***	-44.56	-27.77	148.1***	-67.23	-29.56	2.564	90.06*	106.8**	46.07	17.59	17.49**
R-squared	0.303	0.362	0.268	0.488	0.201	0.228	0.330	0.355	0.431	0.244	0.301	0.286

# Prolongements

---

Introduction

Méthodologie

Présentation  
des données

Quelques  
résultats

*Prolongements*

Tester l'ontologie : identification des auteurs  
« particuliers » qui tendent à utiliser des  
termes spécifiques

Avant tout Finalisation de la BD

⇒ **Fin de la numérisation**

⇒ **Géospatialisation des données**

⇒ **Utilisation de données externes**

⇒ D'occupation du sol

⇒ Régionales et départementales

⇒ Autres documents de description des paysages

# MERCI

---

## Congrès AFS 2014, RT Méthodes, 05.09.2013

*Julie Bourbeillon, UMR IRHS*

*Jonathan Charles, UMR GRANEM*

*Damien Rousselière, UMR GRANEM*

*Julien Salanié, UMR GATE*

*Djery Sow, UMR GRANEM*



**PDD 2**  
Paysage et développement durable