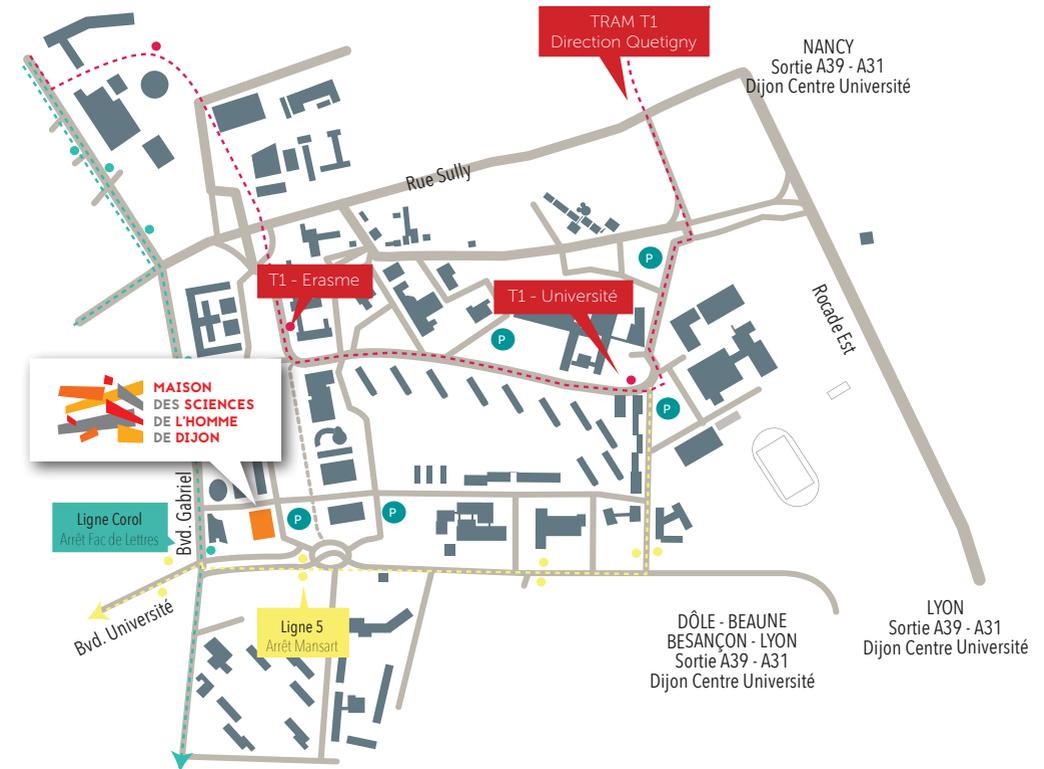


Du 9 au 13 décembre 2019

SEM-AINE DATA/SHS

Maison des Sciences de l'Homme de Dijon

La semaine DATA SHS est un événement composé de 4 conférences quotidiennes faisant intervenir des chercheurs en SHS du grand campus dijonnais ainsi que des partenaires professionnels (Banque de France, INSEE) autour de travaux de recherche mobilisant des données et des méthodes quantitatives.



Programme

Lundi 09/12		Mardi 10/12		Mercredi 11/12		Jeudi 12/12		Vendredi 13/12	
8h30 9h00	accueil café ouverture - Jean VIGREUX, Directeur MSH Dijon - Jimmy LOPEZ, Responsable scientifique PUDD	9h-10h30 Cathy PERRET -CIPE - laboratoire IREDU « Disparition d'un dispositif institutionnel de réorientation : quels effets sur la persévérance universitaire des étudiants de santé ? »		9h-10h30 Inès ALBANDEA - laboratoire IREDU « La méthode des vignettes pour analyser la perception des recruteurs sur les parcours d'études non linéaires »		9h-10h30 Gelareh TAHAMI - laboratoire Psy-DREPI « Image d'un coach idéal dans le sport de haut niveau »		9h-10h30 John A. LEE - Université Catholique de Louvain « Réduction de dimension : d'hier à aujourd'hui, de linéaire à non-linéaire, de petites à grosses données »	
<i>Pause</i>					<i>Pause</i>				
10h-12h15	Banque de France (BDF) - François Bavay (Directeur régional BFC) : « Données; « Importance dans la connaissance, leur fiabilité et la possibilité de tirer des données des enseignements fiables, le rôle de la Banque de France »	10h45-12h15 Edith SALES-WUILLEMIN - laboratoire Psy-DREPI « La Qualité de vie au travail perçue, l'échelle de mesure QUALTRA-Scale »		10h45-12h15 Saliou DIEDHIOU, Jimmy LOPEZ - laboratoire LEDI « Modèle CDM Recherche et développement, éco-innovation et productivité (Enquêtes innovation du CASD) »		10h45-12h15 Adrien BONNACHE, Kenneth J. SMITH - laboratoire CREGO « Comment les stresseurs influencent-ils la performance des comptables ? Equations structurelles méta-analytiques »		10h45-12h INSEE Bourgogne Franche-Comté Charles PILARSKI (Division Conseil, Etudes et Partenariats) « Les transformations territoriales en Bourgogne-Franche-Comté vues au travers des données de l'Insee. »	
								12h-13h Synthèse/clôture	
<i>Déjeuner</i>					<i>Déjeuner</i>				
14h-15h30	Banque de France (BDF) - Renaud LACROIX (Chief data officer) « Présentation des ressources de la Banque de France et de leur accessibilité aux chercheurs »	14h-15h30 Dominique ANDOLFATTO - laboratoire CREDESPO & Dominique LABBE - IEP de Grenoble « Les élections professionnelles en France »		14h-15h30 Stéphanie CHAMBARON-GINHAC - laboratoire CSGA/INRA « Conscience, cognition et choix alimentaires : de la psychologie aux sciences de l'alimentation »		14h-15h30 Lucile PILLOT - MSH Dijon/ plateforme GEOBFC, Thomas ROY - laboratoire ARTeHIS « Rémunération du travail/ données spatiales à partir de ressources historiques »			
<i>Pause</i>					<i>Pause</i>				
15h45-17h30	Maryse GAIMARD, Viginie DEJOUX, David VALAGEAS - Laboratoire CGC « L'évolution du parc et des caractéristiques des logements en France »	15h45-17h30 Roman DELMOTTE - laboratoire CESAER « L'approvisionnement des villes en légumes : le cas de la ville d'Hanoi, Vietnam »		15h45-17h30 Jean-Philippe ANTONI - laboratoire THEMA « Population synthétique : problème de données et géographie quantitative »		15h45-17h30 Laurent GAUTIER - laboratoire TIL « Regarder les arbres derrière la forêt ? Ou Pourquoi le data turn implique-t-il un semantic turn ? »			

Lundi 09/12

8h30 Accueil café
9h Ouverture (Jean VIGREUX & Jimmy LOPEZ)

10h-12h15 **Banque de France - François BAVAY – Directeur régional Bourgogne Franche-Comté**
« *Importance dans la connaissance, leur fiabilité et la possibilité de tirer des données des enseignements fiables, le rôle de la Banque de France* »

Déjeuner

14h-15h30 **Banque de France - Renaud LACROIX - Chief data officer**
« *Présentation des ressources de la Banque de France et de leur accessibilité aux chercheurs* »

Pause

15h45-17h30 « *L'évolution du parc et des caractéristiques des logements en France métropolitaine depuis 1968. Aspects méthodologiques et descriptifs* »
Maryse GAIMARD, Virginie DEJOUX, David VALAGEAS – Laboratoire CGC - Centre Georges Chevrier (UMR Université de Bourgogne-CNRS 7366).

Depuis les années 1960, le parc de logements en France métropolitaine a fortement augmenté et les conditions de logements se sont profondément modifiées. Si des statisticiens de l'INSEE utilisent régulièrement les données issues du recensement et de l'enquête Logement pour décrire ce parc et parfois son évolution (voir en particulier : Alain JACQUOT, 2006, « *Cinquante ans d'évolution des conditions de logement des ménages* », Données sociales, la société française, pp. 467-473), ces analyses n'examinent pas ou rarement les échelons infranationaux. Nous avons donc cherché à décrire ces phénomènes selon les deux dimensions temporelle et territoriale. L'exploitation de ces bases de données dans cette optique soulève toutefois plusieurs difficultés liées aux ruptures de séries et à l'évolution des zonages analytiques de l'INSEE (voir notamment la première partie de : François CUSIN, Hugo LEFEBVRE et Thomas SIGARD, 2016, « *La question périurbaine. Enquête sur la croissance et la diversité des espaces périphériques* », Revue française de sociologie, vol. 57, pp. 641-679). La représentation cartographique de ces données se heurte aux mêmes difficultés.

Nous présenterons les problèmes méthodologiques et théoriques liés à la représentation d'évolutions territoriales, puis mettrons en valeur le gain représenté par l'automatisation des traitements cartographiques grâce à un logiciel tel que R. Nous présenterons enfin quelques points remarquables tels que l'accroissement apparent constaté du parc de logements anciens entre deux recensements.

- **Source des données utilisées :**
 - Recensement INSEE 1968 à 2015 (fichiers détail logements)
 - Enquête Logement INSEE 2013.
 - IGN GEOFLA 2016.
- **Type d'enquêtes :** quantitatives
- **Méthodologie d'analyse :** statistique et cartographie descriptives.

Mardi 10/12

9h-10h30

« *Disparition d'un dispositif institutionnel de réorientation : quels effets sur la persévérance universitaire des étudiants de santé ?* »

Cathy PERRET – Centre d'Innovation Pédagogique et d'Évaluation (CIPE) de l'Université de Bourgogne et laboratoire IREDU – Institut de recherche sur l'Éducation (Université de Bourgogne - EA 7318).

L'échec en premier cycle universitaire s'est constitué en question politique et scientifique, avec la situation spécifique des études de santé. Peu de travaux sont consacrés aux effets des dispositifs institutionnels de réorientation proposés aux étudiants à l'université. Cette recherche vise à identifier la persévérance universitaire des étudiants, selon qu'un dispositif institutionnel de réorientation soit proposé ou non dans un contexte où les étudiants sont exclus des études universitaires de santé compte tenu de leurs premiers résultats. L'étude mobilise des analyses statistiques réalisées à partir de données administratives d'une université pluridisciplinaire française. Les résultats mettent en évidence des effets différenciés de la présence ou de l'absence du dispositif sur la persévérance des étudiants selon les profils des étudiants.

- **Source des données utilisées :** données administratives d'une université : système de gestion des inscriptions et des examens (APOGEE)
- **Type d'enquêtes :** quantitatives
- **Méthodologie d'analyse :** une série de modélisations logistiques.

Pause

10h45-12h15

« *La Qualité de vie au travail perçue, l'échelle de mesure QUALTRA-Scale* »

Edith SALES-WUILLEMIN – Laboratoire Psy-DREPI - PSYchologie : Dynamiques Relationnelles Et Processus Identitaires (Université de Bourgogne – EA 7458).

Les études en psychologie du travail montrent l'impact de variables psychologiques comme le sens attribué au travail ou le sentiment de responsabilité sur la motivation, l'implication ou l'engagement au travail (e.g. Hackman et al. 1976 ; Gospel, 2003 ; Desrumaux, et al. 2011, 2012 ; Collange et al. 2013). Si des recherches montrent l'impact de la QVT perçue (Trist, 1978 ; Bergeron et al. 1986 ; Sirgy et al. 2012), aucune ne montre l'impact de l'écart entre la QVT perçue et idéale sur la satisfaction. C'est l'objectif de notre étude.

Le dispositif repose sur un questionnaire structuré en 6 dimensions décomposés en 37 items (Environnement de travail, Tâches et aménagement du temps de travail, liens entre vie professionnelle et vie personnelle, relations sociales au travail, reconnaissance et soutien organisationnel, évolution professionnelle). Le questionnaire a été distribué à 185 salariés (112 Femmes et 73 Hommes, âge moyen = 36,38 ans). La première partie introduisait la VI: écart entre la QVT perçue et idéale. Deux consignes différentes étaient utilisées pour les 37 items (« Dans votre travail actuel, vous avez : ... » « En général, à quel point jugez-vous inutile/utile pour avoir une bonne qualité de vie au travail »). Les réponses étaient mesurées grâce à une échelle de type Lickert en 5 points allant de tout à fait en désaccord (vs. inutile) à tout à fait d'accord (vs. utile). Un score d'écart entre ces deux mesures était calculé par participant. La deuxième partie du questionnaire introduisait la VD : grâce à un item « mes conditions de travail sont satisfaisantes » et une échelle en 5 points allant de « pas du tout d'accord » à « tout à fait d'accord ». Résultats : L'alpha de Cronbach réalisé sur les items de QVT perçue est de $\alpha = .78$ et de QVT idéale est de $\alpha = .89$. L'analyse des réponses faites au moyen d'une régression linéaire multiple montre que plus l'écart entre la QVT « Idéale » et la QVT « perçue » est élevé, plus le score de satisfaction est faible ($p < .01$). L'étude montre également un impact du statut (cadre).

Les résultats sont discutés en regard de l'intérêt de la prise en compte du point de référence des salariés dans la mesure de la QVT et de l'impact sur leur satisfaction au travail.

- **Source des données utilisées :** questionnaire mis en ligne sur la Plateforme Qualtrics.
- **Type d'enquêtes :** quantitatives
- **Méthodologie d'analyse :** analyse statistiques (Cronbach, régressions linéaires, Analyse de Variance ...).

Déjeuner

14h-15h30 « Les élections professionnelles en France »

Dominique ANDOLFATTO – Laboratoire CREDESPO - Centre de Recherche et d'Etude en Droit et Science Politique (Université de Bourgogne - EA 417) et Dominique LABBE, Institut d'études politiques de Grenoble (IEP de Grenoble).

Depuis 35 ans, nous recueillons régulièrement les séries de données concernant les élections sociales, prud'homales et professionnelles en France, ces dernières à la fois dans le secteur privé et les fonctions publiques. Récemment, nous avons collecté les résultats intégraux de la mesure de la représentativité syndicale telle que celle-ci a été redéfinie en 2008 – pour le secteur privé – et comptabilisé à deux reprises depuis lors, pour les périodes 2009-2012 puis 2013-2016 et avons commencé leur exploitation.

Ces différentes données permettent de retracer l'implantation et l'audience des syndicats dans la société française depuis la Libération, non seulement nationalement mais aussi régionalement et par secteur d'activité.

Au-delà de cet état des lieux, elles ont servi de base pour déterminer les effectifs de chaque organisation syndicale, nationalement et par branche d'activité (fondant les études du ministère du Travail sur le sujet). Rapprochées de la population active, ces séries donnent le niveau de la syndicalisation en France, son évolution depuis 1945 et révèlent les principaux facteurs qui la déterminent.

Il s'agira de présenter ces collections de données, d'envisager leur mise en ligne et de poursuivre la valorisation de leur exploitation.

- **Source des données utilisées :** secteurs public et privé
- **Type d'enquêtes :** élections sociales, prud'homales et professionnelles en France, données sur les résultats intégraux de la mesure de la représentativité syndicale – secteur privé- 2008, 2009-2012, 2013-2016
- **Méthodologie d'analyse :** analyse descriptive

Pause

15h45-17h30 « L'approvisionnement des villes en légumes : le cas de la ville d'Hanoi, Vietnam »

Roman DELMOTTE – Laboratoire CESAER - Centre d'Economie et de Sociologie appliquées à l'Agriculture et aux Espaces Ruraux (AgroSup Dijon – INRA – UMR 1041).

Cette présentation se concentrera sur l'analyse du système alimentaire de la ville d'Hanoi (7 millions d'habitants) à travers son système de distribution, les stratégies d'approvisionnement des acteurs et une estimation de l'empreinte territoriale du système alimentaire. Dans le cadre du projet Food2C (Food to Cities : Dijon, Hanoi, Hue), la méthodologie mixte entre analyse quantitative et qualitative permet d'appréhender les différents aspects (production, consommation, distribution) avec des sources de données variées. Cependant, la variété des sources de données crée des contraintes pour l'analyse scientifique (comptabilité, problèmes d'échelle d'analyse notamment). Ainsi une grande partie de la méthodologie sera concentrée sur comment traiter les différents biais d'analyse possibles. En conclusion, les éléments pouvant être potentiellement réutilisés sur le plan méthodologique seront soulignés pour la comparaison avec d'autres cas d'études.

- **Source des données utilisées :**
 - Recensement de données primaires (données quantitatives)
 - VHLSS (Viet Nam Household Living Standards Survey 2012-2015)
 - Entretiens semi-directifs (données qualitatives)
- **Type d'enquêtes :** enquêtes qualitatives et quantitatives
- **Méthodologie d'analyse :** analyse spatiale de la répartition des marchés, analyse quantitative des vendeurs de légumes (répartition, sources d'approvisionnement et estimations des volumes vendus). Analyse qualitative des résultats des enquêtes auprès des vendeurs. Estimation de la demande urbaine d'Hanoi pour les légumes.

Mercredi 11/12

9h-10h30 « La méthode des vignettes pour analyser la perception des recruteurs sur les parcours d'études non linéaires »

Inès ALBANDEA - IREDU – Institut de recherche sur l'Education (Université de Bourgogne - EA 7318).

À partir d'une méthodologie appelée méthode des vignettes, nous proposons à des recruteurs d'évaluer des CV fictifs en fonction de plusieurs variables. Ils attribuent un score à chaque CV évalué en fonction de la probabilité qu'ils proposent un entretien d'embauche aux candidats. On s'intéresse plus particulièrement dans cette étude aux parcours d'études non linéaires. Les jeunes diplômés fictifs peuvent avoir effectué un master en 5 ans ou avoir connu une interruption temporaire lors de leur parcours d'études (voyage à l'étranger, service civique ou raison non précisée sur le CV). On montre qu'il est essentiel que le candidat précise l'expérience qu'il a vécue lors de son interruption d'études s'il ne veut pas voir ses chances de passer la première phase d'embauche diminuées. Mais interrompre ses études pour voyager à l'étranger apparaît parfois comme un avantage sur le marché du travail.

- **Source des données utilisées :** données récoltées auprès de recruteurs et employeurs
- **Type d'enquêtes :** quantitatives
- **Méthodologie d'analyse :** multiniveau

Pause

10h45-12h15 « Modèle CDM Recherche et développement, éco-innovation et productivité (Enquêtes innovation du CASD) »

Saliou DIEDHIOU et Jimmy LOPEZ – LEDI - Laboratoire d'Economie de Dijon (Université de Bourgogne – EA 7467)

Face à l'intérêt croissant de la communauté internationale de préserver l'environnement, de nombreuses études se sont concentrées sur la question des déterminants des éco-innovations. Dans ce papier, nous étudions ces déterminants sur des données d'entreprises françaises grâce aux modules environnementaux des enquêtes « innovation », aussi appelées CIS pour (« Community Innovation Survey »), des années 2008 et 2014. Nous observons un effet des réglementations environnementales, mais aussi des caractéristiques des entreprises et des marchés sur lesquels elles évoluent. Le modèle structurel que nous utilisons nous permet aussi de déterminer que : (i) ces effets sur les éco-innovations passent par les efforts de R&D des entreprises et (ii) les éco-innovations permettent d'accroître la productivité des entreprises.

- **Source des données utilisées :** données des enquêtes CIS (Community innovation survey)
- **Type d'enquêtes :** quantitatives
- **Méthodologie d'analyse :** Modèle CDM (Crepon-Duguet-Mairesse)

Déjeuner

14h-15h30 « *Conscience, cognition et choix alimentaires : de la psychologie aux sciences de l'alimentation* »

Stéphanie CHAMBARON-GINHAC – CSGA - Centre des Sciences du Goût et de l'Alimentation (AgroSup Dijon – CNRS – INRA - Université Bourgogne Franche-Comté – UMR A1324 - UB 80).

Il est maintenant bien établi en psychologie qu'une part significative des comportements et des choix sont influencés par des processus non-conscients. Les comportements alimentaires n'échappent pas à cette règle. L'objectif de mes travaux de recherche consiste à adapter et utiliser des méthodologies et les avancées issues de la psychologie cognitive pour étudier les comportements alimentaires. A cette fin, un paradigme d'amorçage a été adapté et utilisé. L'amorçage repose sur le fait que la perception d'un stimulus (l'amorce), qu'elle soit consciente ou non, peut entraîner une modification du traitement d'un autre stimulus (la cible), et aboutir à une modification du comportement. Évolutivement et fonctionnellement étroitement liée à l'alimentation, l'olfaction s'avère être une modalité particulièrement intéressante pour l'étude d'effets d'amorçage implicites sur des comportements alimentaires.

Au travers de cet exposé, je vous présenterai différentes recherches mises place pour étudier les effets d'odeurs alimentaires sur des choix alimentaires et des consommations dans différentes populations (adultes, enfants, personnes âgées) et de différents statuts pondéraux.

Nos travaux démontrent : (1) qu'une odeur de fruit non-consciemment perçue peut avoir un impact sur les intentions de choix et sur les choix en situation réelle de consommation chez des adultes « sains », les guidant vers plus de fruits ou légumes ; (2) qu'une odeur de viande diffusée peu de temps avant le déjeuner semblerait pouvoir augmenter la consommation du plat principal chez des personnes âgées atteintes de la maladie d'Alzheimer et (3) que des odeurs fruitées ou grasses sucrées impacteraient différemment les choix alimentaires selon le statut pondéral des individus.

- **Source des données utilisées** : études expérimentales auprès d'hommes et de femmes en laboratoire
- **Type d'enquêtes** : quantitatives
- **Méthodologie d'analyse** : modèles statistiques

Pause

15h45-17h30 « *Population synthétique : problème de données et géographie quantitative* »

Jean-Philippe ANTONI – Laboratoire THEMA – Laboratoire de géographie (Université de Bourgogne-CNRS - UMR 6049).

Les nouvelles approches de la modélisation urbaine, telles que les systèmes multi-agents (MAS) ou les modèles basés sur les activités (ABM), nécessitent l'utilisation de données individuelles désagrégées. Mais pour des raisons de protection de la vie privée, ces données sont souvent difficiles à obtenir. Une façon de contourner cet obstacle consiste à générer une population synthétique. La méthode présentée dans cette communication à partir de l'exemple de Besançon est subdivisée en deux étapes : (1) une population d'agents appartenant à des ménages attribués à des logements est générée à partir des données socio-démographiques ; (2) Cette population est localisée dans les bâtiments générés à partir des données géographiques. La méthode produit des résultats utiles mais aussi des difficultés pour certaines catégories de population. Au final, elle offre une image tridimensionnelle réaliste de données socio-démographiques individuelles, qui peuvent être cartographiées et utilisées pour modéliser les dynamiques urbaines par l'intermédiaire de MAS ou d'ABM.

- **Source des données utilisées** : INSEE, IGN
- **Type d'enquêtes** : quantitatives
- **Méthodologie d'analyse** : quantitatives

Jeudi 12/12

9h-10h30

« *Image d'un coach idéal dans le sport de haut niveau* »

Gelareh TAHAMI – Laboratoire Psy-DREPI - PSYchologie : Dynamiques Relationnelles Et Processus Identitaires (Université de Bourgogne – EA 7458).

Ce travail s'intéresse à l'alliance des méthodologies qualitatives et quantitatives dans une étude en science humaine. Les spécificités d'un coach idéal identifiées par l'approche qualitative, ont été approfondies par une approche quantitative.

Dans un premier temps, nous avons identifié les spécificités attendues d'un coach idéal, par les athlètes de haut niveau ($n=40$; $M= 23.37$ ans, $ET= 5.27$ ans) dans les sports d'opposition, à travers des analyses thématiques d'entretiens non-directifs. Les quatre spécificités retenues sont: « savoir », « organisation du travail », « expérience » et « comportement et caractère ».

Ensuite, l'approfondissement quantitatif de ces spécificités, nous a permis de vérifier si celles-ci sont différentes en fonction du sexe des athlètes, de la typologie sexuée du sport pratiqué ou des deux à la fois.

Il apparaît une différence significative sur la spécificité « comportement et caractère » entre les joueuses et les joueurs (sport masculin, $p < .01$), différence qui n'apparaît pas entre les joueuses et les joueurs (sport neutre, $p = NS$). En se référant à la théorie de l'identité sociale (TIS), ce résultat montre que les joueuses se servent d'une stratégie de créativité sociale afin de rehausser leur identité sociale menacée.

- **Source des données utilisées** : 40 entretiens non-directifs
- **Type d'enquêtes** : qualitatives
- **Méthodologie d'analyse** : qualitatives

Pause

10h45-12h15

« *Comment les stresseurs influencent-ils la performance des comptables ? Equations structurelles méta-analytiques* »

Adrien BONNACHE et Kenneth J. SMITH – Laboratoire CREGO - Centre de Recherche en Gestion des Organisations (Université de Bourgogne - EA 7317).

Cette recherche vise à combiner les études sur les liens entre les stresseurs et la performance en cabinets comptables et d'audit et à identifier les médiateurs et modérateurs des relations entre stresseurs et performance. En utilisant des méta-analyses et analyses des pistes causales, cette recherche compile 60 études pour investiguer les relations des stresseurs avec la performance des comptables et auditeurs. Comme supposé, les méta-analyses des corrélations bivariées suggèrent que (1) les stresseurs au travail affectent négativement la performance, (2) l'épuisement professionnel et le stress sont reliés négativement à la performance, mais (3) la motivation est associée positivement à la performance. De plus, une modélisation par les équations structurelles méta-analytiques suggère que les tensions au travail ont des effets sur la performance via le stress et l'épuisement professionnel et directs. L'accumulation de plusieurs échantillons par le biais d'une méta-analyse augmente la puissance statistique relativement à des études à échantillon unique et révèle donc le signe de ces effets résiduels directs des tensions de rôle sur la performance en cabinets comptables et d'audit.

- **Source des données utilisées** : 60 études sur les relations des stresseurs avec la performance des comptables et auditeurs
- **Type d'enquêtes** : pas d'enquêtes (résultats de 60 études)
- **Méthodologie d'analyse** : méta-analyse

Déjeuner

14h-15h30 « *Rémunération du travail/données spatiales à partir de ressources historiques* »

Lucile PILLOT – Maison des Sciences de l'Homme de Dijon – plateforme GEOBFC (Université de Bourgogne-CNRS - USR 3516) et Thomas ROY – Laboratoire ARTEHIS - Archéologie, Terre, Histoire, Sociétés (Université de Bourgogne – CNRS UMR 6298).

La rémunération du travail est un concept qui nous est proche ; or, les sociétés médiévales l'ont-ils connu ? Si oui, sous quelles formes et dans quelles proportions ? Voici, les questions centrales qui guident ce travail avec comme point d'amarrage : Dijon à la fin du Moyen Âge. La richesse du patrimoine archivistique médiéval qui concerne la ville de Dijon offre un matériel de première main pour l'étude des rémunérations médiévales. Un premier point d'observation s'est établi sur la période 1370-1395 où le croisement de différents documents permet d'observer structurellement la place des rémunérations au sein de la société dijonnaise. Deux autres points d'observation ont été placés sur les domaines viticoles ducaux et sur la législation municipale en matière de prix et de salaire dans une perspective diachronique élargie au XV^{ème} siècle. La création d'une base de données a permis de recueillir plus de 12000 rémunérations couvrant la période 1370-1395. Elle rend compte d'une activité rémunératrice importante tant dans le secteur du bâtiment, de l'artisanat et de la vigne. Cette Base de données fournit de nombreuses informations qui permettent une exploration spatiale assez fine. Pour ce faire, un important travail de restitution des données spatiales historiques a été effectué par la plateforme GéoBFC de la MSH de Dijon. Il se base sur le recueil des différentes sources d'archives historiques de type cartes et plans, et sur la numérisation des limites anciennes des différentes échelles d'analyse de l'étude (diocèses, archévêchés, doyennés et ville de Dijon –rues et paroisses dijonnaises). Un travail de cartographie et d'analyse des données a été également réalisé, permettant ainsi d'aborder l'analyse historique dans sa dimension spatiale:

- **Source des données utilisées** : données d'archives historiques pour les données thématiques ; données d'archives historiques : cartes et plans.
- **Type d'enquêtes** : quantitatives et qualitatives
- **Méthodologie d'analyse** : approche historique, analyse spatiales et cartographie.

Pause

15h45-17h30 « *Regarder les arbres derrière la forêt ? Ou Pourquoi le data turn implique-t-il un semantic turn ?* »

Laurent GAUTIER – Laboratoire TIL - Centre Interlangues « Texte, Image, Langage » (Université de Bourgogne – EA 4182).

La facilité d'accès aux données textuelles (que ce soit à travers les corpus de référence, les outils de constitution de corpus ad-hoc ou les collectes de données issues des réseaux sociaux) tout comme l'apparente simplicité d'approches strictement statistiques ne sont pas sans poser de nouvelles limites à l'analyse et à l'interprétation des résultats quand les données sont textuelles et ont donc une matérialité langagière. Le goût pour les corpus massifs – dont participe la course au big data – semble justifier le recours à des extractions algorithmiques et des appariements de mots/termes dont on finit par oublier la face signifiante, dans son épaisseur socio-discursive. La question de la comparabilité de toutes les données compilées – considérées depuis leur individualité énonciative – participe aussi largement à ces limites. La présentation visera donc à interroger ces dernières et cherchera à montrer que pour ce type de données, qui mettent en jeu des questions de co-construction et gestion du sens, les paradigmes de référence restent intrinsèquement sémantiques, donc linguistiques.

- **Source des données utilisées** : données textuelles de toute nature : collectées à partir de données txt, produites expérimentalement, collectées sur twitter.
- **Type d'enquêtes** : qualitatives et quantitatives
- **Méthodologie d'analyse** : extractions quantitatives à partir des outils usuels pour les données textuelles (antconc, tkm, sketchengine, ...) pour en montrer les limites et discuter leur dépassement en recourant à la discipline de référence pour l'analyse du sens, la sémantique.

Vendredi 13/12

9h-10h30

« Réduction de dimension : d'hier à aujourd'hui, de linéaire à non-linéaire, de petites à grosses données »

John A. LEE – Institute of Information and Communication Technologies, Electronics and Applied Mathematics - Université Catholique de Louvain (Belgique)

La réduction de dimension a plus d'un siècle d'histoire derrière elle et précède même l'invention des ordinateurs, sans lesquels on ne l'envisagerait même plus aujourd'hui.

Les premières ébauches ont commencé avec des projections linéaires des données, en essayant de préserver la structure de la matrice de covariance ou les dissimilarités entre observations.

Des approches inspirées de la biologie, comme les réseaux de neurones artificiels, ont connu un certain succès, telles les cartes auto-organisées, et tendent à ressurgir à l'ère de l'apprentissage profond.

Pour le moment, cependant, ce sont des stratégies basées sur la préservation des voisinages locaux qui remportent le plus de succès et ont permis des progrès fulgurants il y a quelques 10 ans d'ici. Des méthodes comme t-SNE sont toujours à la pointe, avec des variantes approximées mais largement accélérées, qui permettent d'immerger jusqu'à des millions de points, pour obtenir des représentations très belles et très fines.

Toutes ces méthodes sont applicables à bien des problèmes dans de nombreux domaines, mais leur compréhension reste imparfaite. Avec des indicateurs de performance spécialement conçus, nous essaierons d'y voir plus clair...

- **Source des données utilisées :**

- UCI ML Repository et autres dépôts publics
- Données artificielles

- **Type d'enquêtes :** quantitatives

- **Méthodologie d'analyse :** apprentissage automatique.

Pause

10h45-12h

INSEE Bourgogne Franche-Comté - Charles PILARSKI, Chef de la division Conseil, études et partenariats - « *Les transformations territoriales en Bourgogne-Franche-Comté vues au travers des données de l'Insee* »

L'objectif de la présentation est de mettre en évidence les mutations territoriales depuis une quinzaine d'année sur le périmètre de l'actuelle de Bourgogne-Franche-Comté, au travers de quelques grandes évolutions démographiques, sociales, économiques. En particulier, on s'attachera à dresser une lecture des territoires de la région au travers des dynamiques passées de concentration urbaine, de périurbanisation, et du lien entre espaces ruraux et urbains. Au-delà, l'objectif de la présentation sera également de faire mieux connaître les données localisées produites par l'Insee, qui pour la quasi-totalité sont accessibles pour les chercheurs.

- **Source des données utilisées :** présentation des principales sources de données produites et/ou exploitées à par l'Insee à l'échelon infranational. Le recensement de la population principalement, source qui évolue régulièrement ; mais également les sources sur l'emploi et les entreprises (exploitation des DADS), ou sur les revenus (dispositif Filosofi). Seront également évoquées les sources statistiques qui constituent le futur des données à l'Insee (données de caisse, déclaration sociale nominative, données de téléphonie mobile).

- **Type d'enquêtes :** principalement quantitatives, données administratives.

- **Méthodologie d'analyse :** la présentation mobilisera les résultats de plusieurs études récentes réalisées par l'Insee Bourgogne-Franche-Comté. Elles peuvent selon le thème traité reposer uniquement sur de la statistique descriptive, ou bien des techniques statistiques plus analytiques (analyse de données, régression linéaire).

12h-13h

Synthèse/clôture

Déjeuner