

LES RENCONTRES DE STATISTIQUE APPLIQUÉE

Les outils de la statistique textuelle

Mardi 30 novembre 2010 de 14h à 17h30

Institut National d'Études Démographiques (salle Alfred Sauvy)

- 14h **France Guérin-Pace (Ined) • Introduction de la séance**
- 14h20 **Maryse Marpsat (Insee, Ined et Eris) • Le journal d'un sans-domicile : survie physique et identitaire et évolution** : Dans son journal en ligne *The Panther's Tale*, Albert Vanderburg rend compte presque quotidiennement de sa vie de *homeless*. L'analyse textuelle de ce journal par le logiciel Alceste met en évidence plusieurs mondes lexicaux : ceux associés à la survie physique (se nourrir, dormir, se procurer de l'alcool ou du tabac), dont le vocabulaire indique une posture de témoin, une prise de recul par rapport à ces moments qu'il trouve humiliants ; celui de l'amour et de l'amitié, et ceux qui expriment ses jugements sur l'art et le sens de la vie, où il revendique son droit à vivre sa vie comme il l'entend et à avoir une opinion sur la littérature, la peinture, les événements du monde. Combinée à une analyse de contenu, l'étude de l'évolution de ces thèmes au cours du temps montre une tendance de fond correspondant à l'adaptation à sa nouvelle existence et au relâchement de ses liens antérieurs, mais révèle aussi des remises en question fréquentes de son mode de vie, à la merci des événements fugitifs comme des modifications durables des politiques sociales, de ses ressources ou de sa santé.
- 14h50 **Mathieu Brugidou (Edf r&d Grets, Pacte) et Michèle Moine (Ijk – ms3, Pacte) • Analyse comparative d'énoncés de stigmatisation dans deux questions ouvertes : à propos de normes environnementales émergentes** : Cette communication se propose, à partir de l'analyse de deux questions ouvertes de décrire les formes de stigmatisation suscitées par la transgression de normes environnementales que l'on peut supposer inégalement émergentes : le tri des déchets et les économies d'énergie. Après avoir précisé les conditions statistiques et sociologiques de la comparaison, l'étude montre que la présence d'une norme plus ancienne suscite des discours plus structurés mais aussi plus politisés. La construction d'un lexique de la stigmatisation à partir du vocabulaire de l'émotion permet d'établir que la transgression des règles du tri suscite des stigmatisations plus fréquentes que le refus d'économiser l'énergie. La communication propose enfin d'étudier la compétence normative des membres d'un groupe (reconnaître une déviance et stigmatiser) à partir de formes compétence linguistique. Trois dimensions de cette compétence linguistique peuvent être distinguées : la compétence linguistique -lexicale, syntaxique -, la compétence socio-linguistique et la compétence pragmatique.
- 16h00 **Alexandre Delanoë (Télécom ParisTech – LTCI/CNRS - DEIXIS) • Statistique textuelle et séries chronologiques sur un corpus de presse écrite : le cas de la mise en application du principe de précaution** : Les statistiques textuelles sont opérées sur les articles médiatisant une application du principe de précaution à partir de leur date de publication. Le corpus est extrait de deux bases de données complémentaires de presse nationale et régionale en langue française (Lexis-Nexis et Factiva). L'analyse, réalisée avec plusieurs logiciels (dont Lexico 3 et Tétralogie), révèle que la quantité de formes par article diminue au fur et à mesure que le traitement médiatique s'intensifie dans le temps pour finalement se concentrer sur une terminologie spécifique. La méthode des spécificités chronologiques est mise en œuvre afin d'étudier la dynamique de propagation de la controverse.
- 16h30 **Bénédicte Garnier (Ined) et France Guérin-Pace (Ined) • Les représentations de l'Europe dans le monde à partir d'un corpus de mots associés (projet EuroBroadMap)** : Dans le cadre de ce projet européen, une enquête a été menée auprès d'étudiants (sciences humaines, ingénieurs, art, sciences politiques, etc.) de dix-huit pays (Azerbaïdjan, Belgique, Brésil, Cameroun, Chine, Egypte, France, Hongrie, Inde, Malte, Moldavie, Portugal, Roumanie, Russie, Sénégal, Suède, Tunisie, Turquie), à l'aide d'un questionnaire auto administré. Nous exploitons ici le matériau spécifique constitué des 5100 mots recueillis en réponse à la question « Which words would you associate with Europe ? ». Nous montrerons comment les méthodes de la statistique textuelle (Spad) appliquées à ce corpus permettent de mettre en évidence des représentations de l'Europe très variées, en particulier selon les domaines d'étude et les régions du monde des étudiants interrogés.
- 17h00 **Synthèse et discussion**

Ce séminaire est ouvert à tous, sans frais de participation mais sur inscription à l'adresse suivante :

http://www.ined.fr/fr/rendez_vous/rencontres_statistique_appliquee/

Pour tout autre renseignement contacter Bénédicte Garnier (benedicte.garnier@ined.fr)

Chaque trimestre, le Service Méthodes Statistiques de l'Ined (SMS) propose un séminaire de statistique appliquée. Cette séance autour de la statistique textuelle est organisée avec France Guérin-Pace, responsable de l'unité « Identités et territoires des populations » (UR 12, Ined)