

## Programme séminaire [Agor@ntic](#) « Cooccurrences and co »

**Vendredi 9 janvier 2015**

**Université d'Avignon, campus Hannah Arendt** (centre ville, Ste Marthe, 74, rue Louis pasteur)

<http://www.univ-avignon.fr/fr/informations-utiles/plans-dacces.html>

Salle 0W11 (vieux bâtiment, rez-de-chaussée)

### **MATIN**

8h45 – 9h15 : accueil des participants

9h15 – 9h30 : introduction, tour de table

9h30 – 10h15 : **Carlos Ramisch** (Informatique, LIF-TALEP, AMU)

#### Titre :

Cooccurrence, collocations et expressions : comment trouver une aiguille dans une botte de foin

#### Résumé :

Loin d'être des événements indépendants, l'occurrence des mots dans les textes écrits est gouvernée par des règles complexes de syntaxe, sémantique, entre autres. Cependant, un grand nombre de combinaisons de mots représentent des exceptions à ces règles. Ces combinaisons sont souvent caractérisées par une distribution de cooccurrence non standard. C'est le cas des collocations, comme "pluie battante" et "faire opposition", où les mots ont tendance à s'attirer mutuellement. De manière plus générale, toute combinaison de mots ayant un comportement arbitraire à un certain niveau d'analyse linguistique posera un problème aux applications de TAL. Cette présentation, porte sur le traitement informatique de ces "expressions polylexicales". Je montrerai quelques outils simples pour repérer les expressions et les distinguer de la cooccurrence accidentelle. Ensuite, je discuterai quelques avantages et limitations des approches fondées sur les cooccurrences. Finalement, je présenterai le *lemwetoolkit*, un outil capable d'extraire automatiquement des listes d'expressions à partir des corpus textuels étiquetés.

10h15 – 10h45 : **Marianne Alex** (Sociologie, CNE, UAPV)

#### Titre :

Examiner la masculinité hégémonique : une proposition d'analyse des cooccurrences discursives chez les étudiants

#### Résumé :

L'étude des termes associés au « masculin », relevés par association libre, a été mise en place afin d'approcher les caractéristiques hégémoniques de cette catégorie perçues par les étudiants de l'université d'Avignon. Ceci en tant que groupe social puis de manière individuelle afin d'aborder, entre autres, les « tensions » vécues.

10h45 – 11h : pause café

11h – 11h30 : **Eglantine Schmitt** (Informatique, UTC – COSTECH, Proxem)

#### Titre :

La cooccurrence à l'heure du web : graphes bipartites de catégories linguistiques et de sites web

#### Résumé :

L'exploitation du web pour l'étude de phénomènes sociotechniques et linguistiques est une activité épistémique qui se produit de plus en plus du côté des entreprises (plateformes web, éditeurs de logiciels) et des acteurs individuels (pratique de mesure de soi, de gestion de la réputation personnelle). Cette situation peut être considérée comme une invitation pour les SHS (sciences sociales et sciences du langage notamment) à réimporter dans ces pratiques d'analyse des concepts

et méthodes d'interprétation qui leur sont familières : analyse structurale, analyse de contenus, parcours interprétatif à partir de corrélats statistiques, etc. Ainsi le fait de considérer tout ou partie du web comme un corpus, de même que la nature des unités d'analyse dudit corpus, soulèvent des questions théoriques et méthodologiques peu abordées par ces acteurs. Dans cette perspective, je propose un partage de mes pratiques et réflexions autour d'une expérimentation menée à l'aide du logiciel de collecte et d'analyse sémantique Ubiq, développé par la société Proxem, et consistant à combiner un réseau de cooccurrences de concepts au sein d'un corpus web, avec le réseau des liens de citation des sites mobilisés dans le corpus.

11h30-12h15 : **Frédéric Monier** (Histoire, Université d'Avignon)

Titre :

Prosopographie et analyse de réseaux en histoire : quelques tendances récentes des recherches.

Résumé :

À la différence de la linguistique, de la géographie, ou de la psychologie clinique, la notion de cooccurrence ne fait pas partie des catégories analytiques employées en histoire. Cette distance intellectuelle est d'autant plus frappante que la discipline historique connaît, ces dernières années, un nouvel engouement pour l'usage de méthodes quantitatives d'analyse. Celles-ci proposent un traitement quantitatif des corrélations pour parvenir à identifier des causalités. En ce sens, les cooccurrences seraient en train de devenir une sorte de préalable au travail de l'historien adepte de ces méthodes. Cela est particulièrement vrai dans des analyses de réseaux et dans des travaux de prosopographie. On donnera deux exemples concrets de ces tendances scientifiques, pris dans deux thèses récentes d'histoire contemporaine : Claire Marynowier, Être socialiste dans l'Algérie coloniale. Pratiques, cultures et identités d'un milieu partisan dans le département d'Oran, 1919-1939, Paris, IEP, 2013 ; et Celia Keren, L'évacuation et l'accueil des enfants espagnols en France : cartographie d'une mobilisation transnationale, 1936-1940, Paris, EHESS, 2014.

12h – 12h15 : Discussion

12h30 – 14h : Repas

## APRES MIDI

14h – 14h45 : **Mauro Gaio** (Géomatique, LIUPPA, Université de Pau)

Titre :

La cooccurrence comme méthode de désambiguïsation géographique

14h45 – 15h15 : **Collectif UMR ESPACE** (Géographie, Université d'Avignon, AMU, UNSA)

Titre :

La cooccurrence vue sous l'angle de la géographie et de l'analyse spatiale par les chercheurs de l'UMR ESPACE.

Résumé :

La géographie quantitative manipule fréquemment des notions de cooccurrence ou de corrélation spatiales, voire temporelles. A travers ses pratiques scientifiques et les acceptions de ces termes, le collectif d'ESPACE propose sa vision des cooccurrences et décrit les méthodes et les outils de mesure et d'estimation qui y sont associés.

15h15 – 16h : **Marc El Bèze** (Informatique, LIA, UAPV)

Titre :

Quelques réflexions sur les cooccurrences et les collocations dans des dialogues, des tweets ou des blogs

16h – 17h : Discussion et perspectives