

**mardi 20 mars 2012,  
de 15h à 17h, en amphi K**

## **Max Silberztein**

Professeur de linguistique à l'INALCO et à l'Université de Franche-Comté, présentera une conférence sur le thème :

### ***NooJ : un moteur de recherche pour l'analyse de corpus***

Les moteurs de recherche traditionnels (Google, Bing etc.) permettent de trouver dans le texte des sites WEB des formes orthographiques (c'est-à-dire des chaînes de caractères) mais ne permettent pas de localiser des termes ou des concepts. Certains outils d'exploration de corpus (Frantext, British National Corpus etc.) insèrent des informations linguistiques dans des textes pré-sélectionnés sous forme d'étiquettes ("tags"), ce qui permet aux utilisateurs de rechercher des séquences comme "Déterminant Nom Adjectif Verbe", mais ces étiquettes, fixées une fois pour toutes, ne correspondent pas forcément aux besoins des utilisateurs de corpus en sciences humaines : enseignants de langues, historiens, littéraires, sociologues et psychologues.



Grâce au logiciel NooJ, à l'instar des moteurs de recherche traditionnels, il est possible d'explorer des corpus ouverts (donc non étiquetés au préalable) tout en offrant aux utilisateurs de concevoir leurs propres jeux d'étiquettes. Par exemple, un psychologue peut, au moyen de NooJ retrouver des séquences du type "NomHumain VerbePsychologique NomDeSentiment". Un sociologue pourra étudier, dans un corpus de textes journalistiques, les co-occurrences entre des références à certaines personnalités politiques et l'utilisation d'adverbes positifs (ex.généreusement) ou négatifs (ex. imprudemment), etc.

NooJ fournit également à ses utilisateurs des outils qui leur permettent de définir leurs propres termes, concepts et relations (ou prédicats). Je présenterai ces outils et ferai la démonstration d'une de leurs applications : l'analyse sémantique fine d'un corpus extrait du *Monde Diplomatique* en posant le problème de l'ambiguïté sémantique.

***Amphi K***  
***Centre Pierre Mendès-France***

**90 rue de Tolbiac 75013 Paris, rez-de-chaussée, métro Olympiades.**