

Semaine de formation intensive NooJ
Max Silberztein, Université de Franche-Comté
Auditorium Dumézil ou salle 001 (à vérifier)
INALCO, 2 rue de Lille, 75007 Paris
6-10 janvier 2025

NooJ est un environnement de développement utilisé pour décrire formellement les langues et analyser des textes à l'aide de ressources linguistiques à plusieurs niveaux : orthographe et typographie, morphologie flexionnelle, dérivationnelle et agglutinative, syntaxe locale, structurelle (en constituants) et de dépendance, grammaire transformationnelle et analyse sémantique. NooJ offre tous les outils formels de la hiérarchie de Chomsky-Schützenberger pour formaliser chaque type de phénomène : grammaires régulières (comme XFST), hors-contexte (comme GPSG), contextuelles (comme LFG) et non restreintes (comme HPSG), ainsi que des outils de développement pour créer, tester et partager des ressources linguistiques à large couverture.

NooJ est utilisé comme outil de linguistique de corpus et dans les humanités numériques pour effectuer des analyses statistiques sur des corpus historiques, littéraires, psychologiques ou socio-linguistiques en analyse de discours, cf. <http://atishs.univ-fcomte.fr>. Le moteur de NooJ a été intégré à de nombreuses applications de TAL : reconnaissance d'entités nommées, génération automatique de textes, informatique décisionnelle, interface en langue naturelle, moteur de recherche conceptuel, production automatique de paraphrases, web sémantique, traduction automatique, cf. plus de 500 références à la page nooj.univ-fcomte.fr/references.html.

NooJ est soutenu depuis 2013 par le programme Metashare de la Communauté européenne, est gratuit et open source et fonctionne sous Windows (.NET), Mac OSX, LINUX et Unix (JAVA). Il existe des modules de ressources linguistiques au format NooJ pour plus de 30 langues, cf. : nooj.univ-fcomte.fr/resources.html. Son nouveau moteur linguistique tourne en natif sur Linux, MacOS et Windows, dispose d'une WEB API, et est utilisé pour la version WEB de NooJ <https://webnooj.univ-fcomte.fr>.

Ce stage s'adresse particulièrement aux étudiants de Master, doctorants et collègues universitaires intéressés par la formalisation linguistique, les outils d'exploration de corpus et les applications du TAL. Les étudiants en Master qui peuvent et désirent valider le stage auprès de leur université doivent me prévenir dès leur inscription. Le stage dure une semaine : les matins sont dédiés au cours ; pendant les après-midis, des chercheurs présenteront diverses applications et organiseront des ateliers avec NooJ.

ATTENTION : les places sont limitées. Les inscriptions sont obligatoires : il faut s'inscrire en envoyant son nom, son statut (étudiant / doctorant / universitaire / autre) et le nom de son institution à max.silberztein@univ-fcomte.fr.

Lundi 6 janvier

9H-12H30 : **Corpus** : Limites des approches empiriques et IA. Corpus, requêtes et concordances.
14H-15H : Masako Watabe (Université de Franche-Comté) : Formalisation des adjectifs du romani.
15H-16H : Marzieh Rabiei (Université de Franche-Comté) : Formalisation des verbes du persan.

Mardi 7 janvier

9H-12H30 : **Lexique & morphologie** : Unités Linguistiques Atomiques, dictionnaires et morphologie.
14H-15H : Asmaa Kourtin (Université de Kenitra, Maroc) : Modélisation et implémentation des expressions figées
15H-16H : Asmaa Amzali (Université de Kenitra, Maroc) : Analyse syntaxique en arabe à l'aide des tables lexique-grammaire dans NooJ.

Mercredi 8 janvier

9H-12H30 : **Syntaxe locale** : Requêtes syntaxiques, grammaires locales, annotation de textes, entités nommées.
14H-14H40 : Maximiliano Duran (Université de Franche-Comté) : Grammaires syntaxiques pour la négation en quechua.
14H40-15H20 : Jean Mourier (Université de Montpellier) : Repérer les discours de haine : approches lexicales et argumentatives.
15H20-16H : Agata Jackiewicz et Lara Nougaret (Université de Montpellier) : Approcher la notion de commun(s) par l'exploration de corpus écologique.

Jeudi 9 janvier

9H-12H30 : **Syntaxe structurelle** : Grammaires hors-contexte, contraintes d'accord et arbres.
14H-15H : Olena Saint-Joanis (INALCO) : Module ukrainien pour NooJ.
15H-16H : Fabio Meroni (Université de Bergamo, Italie) : Formalisation du Mwotlap.

Vendredi 10 janvier

9H-12H30 : **Sémantique** : Grammaires transformationnelles. WEB sémantique. Génération automatique de textes.
14H-15H : Andrea Rodrigo (Université de Rosario, Argentine) : Treatment of the ambiguities of Argentinian Spanish.
15H-16H : Amira Abbes (Université de Sfax, Tunisie) : Construction d'un entrepôt de données NoSQL orientées documents des réseaux sociaux.