

Semaine de formation intensive NooJ
Max Silberztein, Université de Franche-Comté
Auditorium Dumézil
INALCO, 2 rue de Lille, 75007 Paris
3-9 janvier 2023

NooJ est un environnement de développement linguistique utilisé pour formaliser tous les phénomènes linguistiques à prendre en compte pour effectuer des analyses/générationnelles transformationnelles automatique : orthographe et typographie, morphologie flexionnelle, dérivationnelle et agglutinative, syntaxe locale, structurelle (en constituants) et de dépendance, grammaire transformationnelle et analyse sémantique. NooJ propose des outils formels adaptés à la description de chaque type de phénomène : grammaires rationnelles (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 aussi utilisé comme outil de linguistique de corpus et dans les humanités numériques pour effectuer des analyses statistiques sur des corpus historiques, littéraires, psychologiques ou sociologiques de corpus, cf. <http://atishs.univ-fcomte.fr>. Le moteur linguistique 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 400 références à www.nooj4nlp.org/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. : www.nooj4nlp.org/resources.html. Un nouveau moteur linguistique est en chantier sur gitlab : gitlab.com/Silberz/ra-linguistic-engine et tourne en natif sur Linux, MacOS et Windows.

Ce stage s'adresse particulièrement aux étudiants de Master, doctorants et 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 et aux travaux dirigés ; pendant les après-midis, des chercheurs présenteront diverses applications / tutoriels de NooJ.

ATTENTION : le stage est gratuit mais 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.

Mardi 3

9H-12H30 : **Corpus** : Limites des approches stochastiques. Corpus, requêtes et concordances.

14H-15H : Jacqueline Léon (CNRS UMR 7597) : Histoire des débuts du TAL en France.

15H-16H : Ismahane Kourtin (Univ. De Franche-Comté) : Question-Réponse dans le domaine juridique.

Mercredi 4

9H-12H30 : **Lexique & morphologie** : Unités Linguistiques Atomiques, dictionnaires, morphologie.

14H-15H : Masako Watabe (Univ. de Franche-Comté) : Le dictionnaire de la langue rromani.

15H-16H : Alena Saint-Joanis (Univ. de Franche-Comté) : Grammaires morphologiques de l'ukrainien.

Jeudi 5

9H-12H30 : **Syntaxe locale** : Requêtes syntaxiques, grammaires locales, annotation de textes, entités nommées.

14H-15H : Mario Monteleone (Univ. de Salerne, Italie) : Continuité morphophonémique et discontinuité sémantique.

15H-16H : Maximiliano Duran (Univ. de Franche-Comté) : Génération automatique des nombres en quechua.

Vendredi 6

9H-12H30 : **Syntaxe structurelle** : Grammaires hors contexte, contraintes d'accord et arbres.

14H-15H : Héla Fehri (Univ. de Sfax) : Un jeu pédagogique multilingue : conjugaison et flexion.

15H-16H : Asmaa Amzali (Univ. de Kenitra) : Les tables lexique grammaire des verbes psychologiques arabes.

Lundi 9

9H-12H30 : **Sémantique** : Analyse et génération de paraphrases automatique, WEB sémantique.

14H-15H : Slim Mesfar (Univ. de la Manouba) : Désambiguïsation en arabe utilisant les mots composés.

15H-16H : Hajer Cheikhrouhou (Univ. de Sfax) : Reconnaissance et traduction automatique des noms prédicatifs arabes.