

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
5-9 janvier 2026

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, 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, annotation psychologique, 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 et dispose d'une WEB API qui est utilisée en particulier 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

9h-12h30 : **Corpus** : Limites des approches empiriques et IA. Corpus, requêtes et concordances.

14h-15h : Christian Boitet (Université de Grenoble-Alpes) : Importance de la linguistique pour le TAL.

15h-16h : Andrea Rodrigo (Universidad Nacional de Rosario, Argentine) : Detection of noun clauses in Argentinian Spanish.

Mardi

9h-12h30 : **Lexique & morphologie** : Unités Linguistiques Atomiques, dictionnaires et morphologie.

14h-15h : Nicolas Boffo (Université de Montpellier) : Extraction de relations dans les textes liés au trafic de drogue.

15h-16h : Amel Zayeni (Université de Sfax, Tunisie) : Analyse des interactions des apprenants sur Moodle avec NooJ.

Mercredi

9h-12h30 : **Syntaxe locale** : Requêtes syntaxiques, grammaires locales, annotation de textes, entités nommées.

14h-15h : Marzieh Rabiei (Université de Franche-Comté) : Morphologie des verbes du persan.

15h-16h : Masako Watabe (Université de Franche-Comté) : Désambiguïsation des phrases AN (adjectif-nom) en rromani.

Jeudi

9h-12h30 : **Syntaxe structurelle** : Grammaires hors-contexte, contraintes d'accord et arbres.

14h-15h : Maximiliano Duran (Université de Franche-Comté) : La phrase interrogative quechua et ses transformations.

15h-16h : Olena Saint-Joanis (INALCO) : Module ukrainien pour NooJ.

Vendredi

9h-12h30 : **Sémantique** : Grammaires transformationnelles. WEB sémantique. Génération automatique de textes.

14h-14h40 : Prihantoro, Ekgoan Susanti Utami (Universitas Diponegoro, Indonesia) : Morphological Grammars for Javanese Verbs: Towards a NooJ Javanese Tagger.

14h40-15h20 : Fabio Meroni (Universidad de Vitoria-Gasteiz, Espagne) : Formalisation de la morphophonologie du mwotlap avec NooJ.

15h20-16h : Divna Petkovic, Victor Rabiet (INALCO) : Grammaires NooJ pour les expressions temporelles en ukrainien et serbe.