

Linguistique-Informatique

Sciences du langage

■ formation initiale

La linguistique informatique, au carrefour des sciences humaines et des sciences exactes.

#linguistique computationnelle, #ingénierie linguistique, #TAL

L'UFR Linguistique de l'université Paris Diderot offre depuis plus de 30 ans un cursus de Linguistique-Informatique avec pour objectif de former des spécialistes en «Traitement Automatique de la Langue» (TAL).

Avec le développement continu des nouvelles technologies de l'information et de la communication et l'explosion du numérique, le traitement informatique de données linguistiques occupe une place essentielle, en particulier avec des applications comme la traduction automatique, la recherche d'information, la correction orthographique, la reconnaissance de la parole, etc.

La mise en œuvre de ces applications demande des compétences à la fois en informatique et en linguistique que vous pourrez acquérir en choisissant la spécialité Linguistique-Informatique du master Science du langage. Les débouchés pour des étudiants formés au TAL sont assurés, que ce soit dans le milieu de la recherche (privée ou publique) ou dans le milieu industriel (nombreuses entreprises spécialisées en ingénierie linguistique).

Les enseignements

Un équilibre disciplinaire entre informatique et linguistique (et, de plus en plus, mathématiques) caractérise cette spécialité en science du langage.

En insistant sur des enseignements fondamentaux dans chaque discipline, cette formation vise à donner une réelle double compétence, permettant de participer à la conception et au développement d'applications de TAL, alors que beaucoup de masters comparables visent simplement la formation d'utilisateurs avertis...

Partenariats privilégiés

PME spécialisées : Arisem, Druide Informatique, Eptica-Lingway, Kwaga, Linkfluence, Memodata, Nuance, Proxem, Sinequa, Synapse, Synomia, Systran, Themis, Trooclick, Vecsys, Viavoo, Virtuoz.

Grands groupes : AFP, CEA, EDF, Thalès, TNS-Sofres, Sony.

Première année

Linguistique :

- phonétique
- phonologie
- morphologie
- syntaxe
- sémantique

Informatique :

- algorithmique
- programmation (objet)

TAL et ses fondements mathématiques :

- méthodes probabilistes
- méthodes par apprentissage automatique
- théorie des langages formels
- traduction automatique

Seconde année

TAL avancé :

- analyse syntaxique automatique
- analyse sémantique automatique
- analyse de données pour le TAL
- analyse et génération de discours
- fouille de texte

Cours applicatifs, assurés par des professionnels du domaine (xml, ontologies, opinion mining, sql...)

Stage de 6 mois en entreprise ou en laboratoire de recherche, validé par un rapport/mémoire et une soutenance.

Le cursus de Linguistique-Informatique a été le premier cursus de ce type en France et il reste à ce jour une référence pour les professionnels comme pour les chercheurs du domaine.

Titres requis

La spécialité Linguistique-Informatique s'adresse aux étudiants ayant déjà une idée précise du champ de la linguistique (phonétique, syntaxe, morphologie...), et ayant déjà une compétence concrète en programmation (connaissance d'au moins un langage de programmation générique, bases d'algorithmique). Un bon niveau en mathématiques (probabilités, automates) favorise également leur réussite.

- Licence pluridisciplinaire en sciences humaines / sciences exactes, (CPGE B/L, MIASHS, bi-licence, frontières du vivant...)
- Formation à dominante mathématiques / informatique, avec des compétences en linguistique
- Formation à dominante sciences humaines, avec des compétences en informatique

À Paris Diderot

- Licence 3^e année « linguistique informatique »
- Licence MIASHS parcours MIL (math / info / linguistique)

Compétences visées

Compétences en linguistique fondamentale et expérimentale

- capacité de description formalisée de tous les niveaux (phonétique, morphosyntaxique, sémantique) d'une langue donnée
- connaissances méthodologiques et statistiques permettant d'exploiter des corpus de données langagières
- connaissance des grandes problématiques de la linguistique moderne (modélisation, acquisition, pathologie, typologie...)

Compétences en informatique

- capacité à spécifier et évaluer les aspects algorithmiques de toute tâche de manipulation de données (textuelles ou numériques)
- compétence complète de programmation dans au moins un langage de scripting (python/perl) et un langage objet (java)
- expérience concrète de développement collaboratif

Compétences en TAL

- connaissance des grandes applications de TAL et des méthodes utilisées depuis le début des travaux de ce domaine
- familiarité avec les environnements de développement et les ressources spécifiques du TAL (unitex, wordnet, nltk...)
- capacité à mettre en œuvre des traitements au niveau de l'état de l'art pour les tâches normalisées courantes du TAL (étiquetage morpho-syntaxique, lemmatisation, chunking, résolution d'anaphores, similarité lexicale...)

Après le master

Secteurs d'activité

- Centres de recherche publics et privés
- PME et grandes entreprises intégrant ces technologies dans des logiciels dont l'usage fait maintenant partie de la vie de tous les jours (traitement de texte, moteur de recherche, téléphone portable).

Métiers exercés

- Ingénieur linguiste dans les PME spécialisées du domaine, ou dans des entreprises qui intègrent les technologies du TAL dans des applications (documentation, publication, éditeurs de logiciels...).

Poursuite d'étude en doctorat

- Ecole doctorale Sciences du Langage (Paris Diderot)
- Ecole doctorale Sciences mathématiques de Paris Centre
- Ecole doctorale Frontières du vivant
- Laboratoires d'accueil privilégiés : ALPAGE (UMR INRIA-Paris Diderot), LLF (UMR CNRS-Paris Diderot).

Modalités d'inscriptions

- Sélection sur dossier
Pour une bonne préparation de la candidature, envoyer un cv à : linguistique.informatique@linguist.univ-paris-diderot.fr

U.F.R. Linguistique (UFRL)

Responsable du master 1

Pascal Amsili
pascal.amsili@linguist.univ-paris-diderot.fr

Responsable du master 2

Laurence Danlos
Laurence.Danlos@linguist.univ-paris-diderot.fr

Contacts

Scolarité :
Armelle Collot
scolarite@linguist.univ-paris-diderot.fr

www.univ-paris-diderot.fr
li.linguist.univ-paris-diderot.fr
www.linguist.univ-paris-diderot.fr