

Master LITL



Domaine : Sciences humaines et sociales

Mention : Sciences du langage

Parcours : **Linguistique, Informatique, Technologies du Langage**

Réunion d'informations étudiants de L3

Principaux membres de l'équipe enseignante :
Cécile Fabre (resp.), Lydia-Mai Ho-Dac, Ludovic Tanguy

Plan de la présentation

- Objectifs du master et compétences développées
- Secteurs d'activité et types d'emplois
- Organisation et contenus
- Projets, mémoires
- Intégrer le master

Objectifs du master LITL

- Le master LITL associe **linguistique et informatique** pour former des spécialistes du **traitement automatique des langues (TAL)**
- Le TAL est un domaine qui vise à :
 - créer des outils capables de traiter de façon automatique des données langagières, et d'assister de nombreuses tâches nécessitant l'analyse du langage :
traduire, répondre à des questions, rechercher de l'information, détecter des connaissances nouvelles, des opinions, synthétiser le contenu des documents, faciliter l'exploitation de données volumineuses (rapports, mails), etc.
 - répondre aux besoins aujourd'hui très importants d'accès à l'information et aux connaissances qui sont exprimées par des textes
- A la fois un domaine de recherche et un champ professionnel, à l'importance toujours croissante, lié au développement de méthodes et d'outils issus de l'**intelligence artificielle**

Objectifs du master LITL

Le master LITL comporte en M2 une formation complémentaire à l'**ergonomie cognitive** : prise en compte des besoins utilisateurs et des usages.

Les diplômés du master LITL sont capables de développer des *ressources linguistiques* et de recourir à des *traitements informatiques* pour concevoir des systèmes de TAL

ET

ils sont **sensibilisés à la prise en compte des usages**, des situations concrètes dans lesquels ces dispositifs vont s'intégrer

En M2 : deux modules mutualisés avec le M2 de psychologie ECIT-FH

Environnement



La formation s'appuie sur

- un laboratoire de recherche CNRS-UT2J (**CLLE : Cognition, Langage, Langues, Ergonomie**) associant des chercheurs en linguistique et en psychologie
= laboratoire d'adossement du master
- le tissu industriel toulousain et un réseau professionnel à Toulouse (Airbus, Safety Data, CNES, Coup de Puce, Inbenta, Prometil, Synapse) et en France (Eptica, Fortia financial solutions, Orange labs, Proxem, Syllabs...)
- un réseau d'anciens étudiants

Les secteurs d'activités concernés

- **Le traitement des contenus textuels**
 - Rechercher, analyser, transformer l'information
 - Extraire des connaissances
 - Produire des documents de meilleure qualité, plus faciles à exploiter
- **Dans tous les domaines**
 - Gestion de la communication client
 - Gestion des risques
 - Veille technologique, commerciale, scientifique
 - Gestion de la documentation technique et des connaissances métier
 - Marketing

Exemples d'emplois occupés par des diplômés du master

- *Dans des petites entreprises (startup, cabinet de conseil, société spécialisée en TAL) :*
 - *Data scientist (Hubware, Toulouse)*
 - *Responsable R&S (Coperbee, Toulouse)*
 - *Knowledge manager (Synapse, Toulouse)*
 - *Ingénieure linguiste (Syllabs, Paris)*
 - *NLP Engineer (Maluuba, Canada)*
 - *Linguiste (Inbenta, Toulouse)*
 - *Product Manager (Proxem, Paris)*
 - *NLP specialist (Linguamatics, Cambridge UK)*
- *Dans des grandes entreprises :*
 - *Terminologue (Airbus, Toulouse)*
 - *Ads Quality Associate in French (Google, Dublin)*
- *Dans des organismes publics :*
 - *Doctorante, contrat doctoral, CLLE, Toulouse*
 - *Doctorante, contrat CIFRE, Prometil & CLLE, Toulouse*
 - *Doctorant, financement ANR, ICAR, Lyon*
 - *Ingénieur d'études à l'INRA (Paris)*
 - *Ingénieur d'études au CNRS (Nancy)*
 - *Ingénieur pédagogie et TICE, COMUE de Toulouse*
 - *Responsable Numérique et Observatoire du doctorat, Université de Toulouse*

Nature des activités professionnelles

- Développer des ressources linguistiques pour améliorer un système de TAL
 - Lexiques, grammaires, terminologies, ontologies
 - Traitement de corpus issus de domaines d'activité divers
- Evaluer des systèmes de TAL
 - Analyse de sorties de systèmes, évaluation, diagnostic
 - Recommandations d'amélioration en termes de traitements linguistiques
- Proposer des traitements innovants et adaptés à des besoins variés
 - Adaptation de techniques existantes à un nouveau domaine
 - Intégration de techniques plus riches en réponse à des demandes
- Etre expert linguiste sur toutes les questions liées au langage
 - Personne ressource pour les traitements linguistiques
 - Chef de projet, responsable recherche & développement
 - Polyvalence (suivi du développement des logiciels, tests, contacts clients/fournisseurs/équipes techniques, analyse des besoins...)

Exemples

- **product manager junior chez Proxem (Paris)** : E. travaille sur des projets de réalisation d'agents conversationnels
- **ingénieure linguiste à Besedo (Paris)** : L. travaille sur le développement de ressources et d'outils pour aider le travail de modération de contenus
- **NLP engineer chez Maluuba (Canada)** : B. développe et améliore un système de dialogue oral en français sur les smartphones
- **ingénieure linguiste informaticienne et responsable R&D chez Sharewizme (Lille)** : S. développe et utilise des systèmes informatiques qui analysent en temps réel les réponses des participants pour animer des séminaires
- **terminologue chez Airbus (Toulouse)** : L. inventorie et organise les termes techniques utilisés dans l'entreprise et les utilise pour améliorer l'accès à la documentation
- **ingénieure linguiste chez Inbenta (Toulouse)** : en lien avec les clients, C. développe des ressources langagières spécifiques pour adapter le moteur de recherche à leurs activités
- **linguiste informaticien chez CFH/Safety Data (Toulouse)** : N. y a réalisé son doctorat (convention CIFRE) dans lequel il a proposé un système innovant pour aider les experts en sécurité aérienne à analyser les rapports d'accidents
- **chercheur post-doctorant au CNRS (Paris)** : F. travaillait sur un système qui analyse les forums de santé pour y identifier les effets indésirables des médicaments. Depuis il travaille dans la société **Linguamatics**, en tant que **NLP specialist**, à Cambridge.

Organisation de la formation

M1 300 h + mémoire		
Enseignements spécifiques	TAL et informatique	100h
Méthodologie	Corpus et linguistique outillée, statistiques	50h
Options de linguistique	Syntaxe, phonologie, sémantique, etc.	75h
Projet tutoré	Projet collectif encadré	25h
Langues	Anglais si C1 non acquis	25h
Mémoire	Travail autonome encadré	

Détail des enseignements du M1 LITL

UE801 SLT0801T Mémoire M1 10 ECTS 25h	UE802 SLT0802V Statistiques appliquées au langage 3 ECTS 25h	UE803 Option LiCo 3 ECTS 25h	UE804 SLT0804T Méthodes, ressources et applications du TAL 1 4 ECTS 25h	UE805 SLT0805T Informatique pour le TAL 2 4 ECTS 25h	UE806 SLT0806T Projet TAL 3 ECTS 25h	UE807 LV LANSAD 3 ECTS 25h	150h
UE701 SLT0701T Projet de recherche 4 ECTS	UE702 SLT0702T Informatique pour le TAL 1 6 ECTS 25h	UE703 SLT0703T Méthodes, ressources et applications du TAL 1 6 ECTS 25h	UE704 Option LiCo 4 ECTS 25h	UE705 Option LiCo 4 ECTS 25h	UE706 SLT0706T Corpus et linguistique outillée 3 ECTS 25h	UE707 LV LANSAD 3 ECTS 25h	150h

Organisation de la formation

M2 250h + mémoire ou stage		
Enseignements spécifiques	Programmation et TAL	100h
Ergonomie	TAL et ergonomie, conception informatique et ergonomie	25h ou 50h
Options de linguistique	Syntaxe, phonologie, sémantique, etc.	25h ou 50h
Professionnalisation	Recherche de stage et projet professionnel	25h
Projet tutoré	Projet collectif encadré	25h
Langues	Anglais si C1 non acquis	25h
Mémoire ou stage long	Travail autonome encadré	

Période d'enseignement : sept-février

N.B. démarrage 1 à 2 semaines avant la date de rentrée universitaire UT2J : pré-rentrée, soutenances de stages de la promotion précédente.

Stage de 4 à 6 mois ou mémoire de recherche

Modalités pédagogiques et évaluation

- Beaucoup de travail pratique
- Des évaluations en continu :
 - Exercices d'une séance à l'autre, compte-rendus, fiches de lecture...
- Travail personnel régulier en particulier pour l'apprentissage de la programmation
- Une dimension « promo » importante :
 - Projets collectifs
 - Créneaux d'accès en salle machine pour faciliter le travail en groupe
 - Tutorat des M1 par les M2

Le mémoire

- Une part très importante du travail à fournir dans l'année
 - Travail de longue haleine permettant de mettre en œuvre les compétences acquises
- Un travail visant à répondre à une question de recherche, en combinant :
 - Des traitements informatiques : manipulation de données linguistiques (corpus, lexiques), utilisation d'outils
 - Une analyse linguistique : annotations, descriptions, analyses

Le mémoire

- Sujets de mémoire proposés par l'équipe pédagogique élargie
 - Présentation pendant la réunion de pré-rentree
- Choix du sujet dans les 3 semaines qui suivent la rentrée
- Principes :
 - Des sujets qui s'inscrivent dans des projets de recherche en cours dans lesquels les directeur.ice.s du mémoire sont impliqu.e.s
 - Articulant analyse linguistique et traitement informatisé de données textuelles

Les projets tutorés

- **Projet TAL en M1**
 - Projet en équipe avec accompagnement pédagogique
 - Réalisation d'une chaîne de traitement automatique complète
 - Projets 2015-2016, 2016-2017, 2019-20 : participation à une campagne d'évaluation internationale
 - Extraction des termes médicaux en français
 - Extraction des causes de décès à partir de rapports médicaux en français et en anglais
 - Détection tweets mentionnant des effets secondaires de médicaments
 - Projet 2020-21 : construction et caractérisation d'un corpus de textes procéduraux diffusable et exploitable par la communauté scientifique

Les projets tutorés

- **Projet conception en M2**
 - Projet collectif en réponse à une demande émanant d'un terrain professionnel
 - Analyse de la demande, proposition d'intervention et gestion des différentes étapes de conception d'une chaîne de traitement
 - Restitution écrite et orale au commanditaire
 - Ex de projets :
 - Extraction d'informations à partir de compte-rendus médicaux (CHU Toulouse)
 - Constitution de ressources textuelles et d'exercices pour la didactique du français langue étrangère à partir de corpus de sous-titres (Archean Labs)
 - Evaluation de la lisibilité des articles de l'encyclopédie Vikidia, destinée à des jeunes lecteurs de 8 à 13 ans
 - Aide à la correction d'erreurs pour des apprenants dyslexiques (Synomia)
 - Implémentation d'un Wiki pour favoriser la transmission des connaissances au sein du CNES
 - Faciliter la navigation des utilisateurs du forum « Vivre sans thyroïde »

Intégrer le master LITL 1^{ère} année

- **Critères d'admission**

Le candidat doit :

- être titulaire d'une licence des mentions conseillées ou d'un diplôme de niveau Bac +3 en lien avec le master LITL
- faire état de compétences suffisantes en linguistique
- avoir une très bonne maîtrise de l'expression écrite et orale en français (pour les étudiants non francophones : attestation du niveau C1 minimum)
- avoir une maîtrise suffisante de l'anglais pour lire des articles scientifiques dans cette langue
- posséder des compétences de base en informatique garantissant un premier niveau d'autonomie dans l'utilisation de l'ordinateur (stocker, organiser, traiter ses données, maîtriser les outils de bureautique).

Le jury appréciera également la qualité du dossier universitaire ainsi que la cohérence du parcours et du projet présentés dans la lettre de motivation.

<http://sciences-du-langage.univ-tlse2.fr/>
Rubrique : MASTERS → LITL