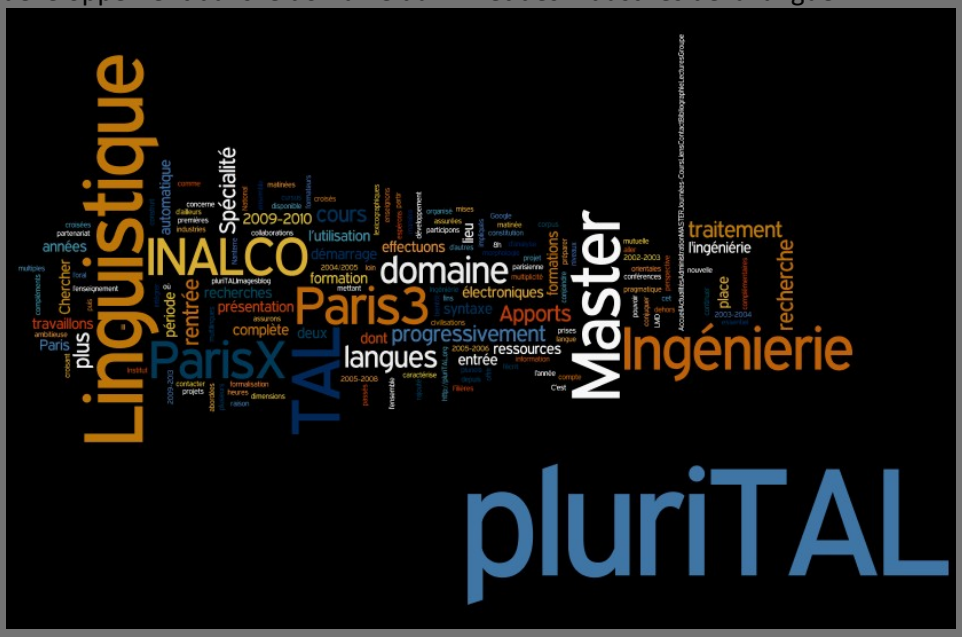
pluriTAL

# MASTER spécialité « Ingénierie Linguistique »

## Université Paris Ouest Nanterre La Défense
## Université Paris III Sorbonne nouvelle
## INALCO

La spécialité « Ingénierie linguistique » concerne la recherche et le développement dans le domaine du TAL et des industries de la langue.




http://plurital.org

# Formation proposée par
## l'Université Paris Ouest Nanterre La Défense
## l'Université Paris III - Sorbonne nouvelle
## l'INALCO

**Domaines :**

***Sciences humaines et sociales (Paris Ouest Nanterre La Défense)***

***Science Humaine et Sociale (Paris III)***

***Langues Cultures et Sociétés du Monde (Inalco)***

**Mentions :**

***Linguistique générale et linguistique appliquée (Paris Ouest Nanterre La Défense)***

***Sciences du Langage (Paris III)***

***Sciences du Langage et Langues appliquées (Inalco)***

**Spécialité :** ***Ingénierie linguistique***

**Responsables:**

Marcel Cori (Paris Ouest Nanterre La Défense)
Serge Fleury (Paris III – Sorbonne nouvelle)
Monique Slodzian (INALCO)

**Commission pédagogique :**

Jean-Michel Daube (INALCO)
Serge Fleury (Paris III – Sorbonne nouvelle)
Marcel Cori (Paris Ouest Nanterre La Défense)

**Hypertoile du MASTER :** http://plurital.org (désormais « le site pluriTAL»)

# Sommaire

Contacts ..... 5
Inscriptions pédagogiques et réunions de rentrée ..... 5
Groupe PluriTAL (liste de diffusion) ..... 5
Localisations ..... 6
Paris Ouest Nanterre La Défense ..... 6
Paris III / ILPGA ..... 6
INALCO ..... 6
Inscriptions – validations d'acquis ..... 7
Inscription en 1ère année ..... 7
Inscription en 2ème année ..... 7
Jury ..... 7
La spécialité « Ingénierie linguistique » ..... 8
Un partenariat universitaire pour le TAL (pluriTAL) ..... 8
Objectifs d'apprentissage ..... 9
Débouchés ..... 10
Organisation globale des enseignements du master ..... 11
Tableau synthétique des différents parcours en M1 et M2 ..... 13
MASTER 1ère année ..... 14
Semestre 1 ..... 14
Semestre 2 ..... 15
Master 1, tous parcours Paris III, Paris Ouest Nanterre La Défense ..... 16
Master 1, INALCO ..... 17
Code des enseignements du M1 (semestre 1) ..... 18
Code des enseignements du M1 (semestre 2) ..... 19
MASTER 2ème année ..... 20
Parcours D : M2 DEFI, Paris Ouest Nanterre La Défense ..... 21
Parcours R : M2 R&D, Paris Ouest Nanterre La Défense, Paris III, Inalco ..... 22
Parcours T : M2 Traductique, Inalco ..... 23
Parcours I : M2 Ingénierie Multilingue, Inalco ..... 24
Contenu des unités d'enseignement ..... 25
Planning des cours du Tronc commun du Master Ingénierie Linguistique ..... 28
Equipe pédagogique ..... 30

Descriptif et horaires des cours (1ère et 2ème années), tronc commun ..... 32
Descriptif et horaires des cours du master 1ère année ..... 32
Grammaires formelles ..... 32
Analyse syntaxique automatique ..... 32
Gestion informatique du multilinguisme ..... 32
Traitement de l'oral / Informatique et phonétique ..... 33
Programmation et projet encadré (semestre 1) ..... 33
Bases de données ..... 33
Statistique et analyse multidimensionnelle ..... 33
Corpus parallèles/comparables ..... 34
Recherche et extraction d'information ..... 34
Syntaxes formelles ..... 34
Syntaxes formelles ..... 34
Document structuré (M1) ..... 35
Programmation et projet encadré (semestre 2) ..... 35
Programmation et algorithmique 1 et 2 ..... 35
Génération automatique de Textes ..... 36
Descriptif et horaires des cours du master 2ème année ..... 37
Formalismes syntaxiques et traitement automatique ..... 37
Document structuré : XML (M2) ..... 37
Analyse du discours et lexicométrie ..... 37
Écritures et multilinguisme ..... 37
Multimodalité du document numérique ..... 38
Sémantique des textes multilingues ..... 38
Modélisation des langues : mots et grammaire ..... 38
Contacts ..... 39

## Contacts

Voir la page « Contact » sur la page web du MASTER.

http://pluriTAL.org/contact.html

## Inscriptions pédagogiques et réunions de rentrée

**Journée d'accueil du MASTER** : (date disponible en ligne sur le site pluriTAL)

**Début des cours :** (date disponible en ligne sur le site pluriTAL)

**Inscriptions pédagogiques :**

| **PARIS III** (*cf* **secrétariat ILPGA ou sur le site pluriTAL**) | **PARIS X** (*cf* **secrétariat Paris X**) | **INALCO** (*cf* **secrétariat INALCO**) |
|---|---|---|

## Groupe PluriTAL (liste de diffusion)

Inscription **obligatoire** pour tous les étudiants devant suivre des cours du Master « Ingénierie Linguistique ». Voir la page « Groupe » sur la page web du MASTER (site pluriTAL).

Lien direct : http://tal.univ-paris3.fr/plurital/groupeplurital.html

Inscription au groupe Yahoo-PluriTAL sur cette page (avec liste de diffusion modérée)

# Localisations

## Paris Ouest Nanterre La Défense

Pour se rendre l'Université de Nanterre : RER A, Direction Saint Germain-en-Laye, station Nanterre Université. Plan de l'université voir http://www.u-paris10.fr. Pour se rendre au bâtiment L. **a)** sortir du RER en queue de train en venant de La Défense/Paris, sortie principale par un escalier montant, prendre sur la gauche, en descendant une rampe d'accès. **b)** en bas de la rampe, tourner immédiatement à droite à 90 degrés. **c)** S'engager à droite dans le passage pratiqué dans la haie une quinzaine de mètres plus loin. **d)** suivre le chemin goudronné qui passe sous le bâtiment (pilotis) **e)** suivre la route parsemée de brise-vitesses, qui longe le bâtiment de l'université à G. et une clôture à D. avec un bâtiment flambant neuf **f)** au bout d'un moment, à travers un cèdre, on distingue un bâtiment aux lignes volontaires. C'est le bâtiment L. Les machines à café sont sur la mezzanine, au premier étage.

## Paris III / ILPGA

ILPGA / Université Paris III Sorbonne nouvelle,
19 rue des Bernardins, 75005 Paris - **Tél.** : 01 44 32 05 70 **Fax** : 01 44 32 05 73
*Plan du quartier* : http://voici.monplan.com/ilpga

## INALCO

Les cours de l'INALCO (M1 et M2) ont lieu au 49bis, avenue de la Belle Gabrielle, 75012 PARIS. RER A, direction Boissy Saint Léger, arrêt à Nogent sur Marne, sortie Marronniers, puis 5 minutes de marche, ou METRO : Château de Vincennes, puis bus 112.

## Inscriptions – validations d'acquis

### Inscription en 1ère année

L' étudiant devra être titulaire d'une licence dans l'une des mentions ou spécialités suivantes : « Sciences du Langage » ; « Lettres » ; « Langues, littératures et civilisations étrangères » ; « Sciences humaines et sociales » ; « Psychologie » ; « Mathématiques appliquées aux sciences sociales » ou d'une bi-licence ou encore d'une licence inter-mentions ayant une composante de Sciences du langage (ex. « Sciences du langage, civilisation européenne : langue » ; « Lettres/sciences du langage »)

### Inscription en 2ème année

L'admission en 2ème année se fait sur dossier y compris pour les étudiants reçus en master 1. Les étudiants titulaires d'un master de Sciences du langage ou de LLCE pouvant faire état d'une appétence pour l'informatique, ou les étudiants en informatique ayant manifesté de l'intérêt pour les langues ou la linguistique verront notamment leur dossier examiné avec intérêt.

Les commissions qui statueront les admissions sont distinctes selon les parcours. Les étudiants peuvent ainsi déposer plusieurs dossiers de demande d'admission. Pour le parcours *Recherche et Développement*, il est demandé aux étudiants d'obtenir l'accord d'un directeur de recherche qui supervisera leur mémoire.

### Formation continue

Le master est aussi ouvert en formation continue pour les traducteurs, documentalistes, bibliothécaires, gestionnaires de sites Web, employés du tertiaire soucieux de se former à des technologies innovantes détenteurs d'une licence ou d'un équivalent par validations d'acquis

### Jury

Un jury se réunit en fin de premier et de second semestre de MASTER pour évaluer les résultats obtenus par les étudiants et faire organiser, le cas échéant, des sessions de rattrapage dans les matières où les étudiants auraient échoué.

## La spécialité « Ingénierie linguistique »

La spécialité *Ingénierie linguistique* concerne la recherche et le développement dans le domaine du TAL et des industries de la langue. L'ingénierie linguistique fait appel à des méthodes et des savoirs multiples. Il s'agit de :

1. Disposer des pré-requis en linguistique : maîtriser les manipulations débouchant sur des descriptions détaillées de faits de langue, connaître les bases des différents domaines des sciences du langage (phonétique et phonologie, morphologie, syntaxe et sémantique) ;
2. Connaître les bases de la recherche et de l'extraction d'information, de la constitution et de la gestion de corpus (écrits ou oraux) et de ressources, y compris multilingues : les corpus sont des mines d'information pour une description réaliste d'emplois d'une langue, les techniques de la recherche et de l'extraction d'information permettent de rapatrier les documents ou les parties de documents jugés pertinents pour un besoin de recherche particulier ;
3. Exprimer les règles et les régularités à l'œuvre dans les corpus, par le biais des grammaires formelles et des traitements quantitatifs pour savoir passer d'une description linguistique d'un texte à une représentation plus formelle permettant sa prise en charge par des logiciels.

L'objectif de la formation est de donner à des étudiants issus des cursus de langues ou de sciences du langage des bases solides qui leur permettent de s'orienter vers les métiers de l'ingénierie linguistique, et de leur donner les possibilités de choisir entre diverses perspectives : document électronique, ingénierie multilingue, traductique. Il s'agit aussi de permettre à certains d'entre eux d'opter pour la recherche et le développement en ce domaine.

## Un partenariat universitaire pour le TAL (pluriTAL)

Le diplôme est délivré par les 3 partenaires suivants :

- Université Paris Ouest Nanterre La Défense
- Université Paris III Sorbonne nouvelle
- Institut National des Langues et Civilisations Orientales (INALCO)

La formation s'appuie sur les laboratoires :

Paris Ouest Nanterre La Défense - **MODYCO** (Modèles, Dynamiques, Corpus, UMR 7114),

http://www.modyco.fr/

Paris III - **SYLED** Centre de Lexicométrie et d'Analyse Automatique des Textes (CLA²T),

http://www.cavi.univ-paris3.fr/ilpga/syled/

Paris III – **LPP** (UMR 7018) Laboratoire de Phonétique et Phonologie

http://lpp.univ-paris3.fr/

INALCO - **ER-TIM** (EAD 2540) : Équipe de Recherche « Textes, Informatique, Multilinguisme »

http://www.crim.fr/

## Objectifs d'apprentissage

**Objectifs d'apprentissage de la spécialité en termes de connaissances (connaissances disciplinaires, connaissances pluridisciplinaires sur l'objet étudié, connaissances méthodologiques, connaissances linguistiques, …)**

- Savoirs disciplinaires en linguistique (en complément de bases solides en phonétique/phonologie, morphologie, syntaxe et sémantique) : sémantique formelle, sémantique lexicale, systèmes d'écriture, traductologie, traductique ;
- Savoirs en TAL : grammaires formelles, syntaxe formelle, analyse syntaxique automatique, gestion du multilinguisme, statistique et analyse multidimensionnelle, traitement de l'oral, recherche et extraction d'information, corpus alignés ;
- Savoirs en informatique : programmation et algorithmique spécifique, bases de données, document structuré (XML) ;
- Maîtrise en réception puis en production de l'anglais scientifique.

**Objectifs d'apprentissage de la spécialité en termes de compétences**

Savoir s'intégrer dans un projet collectif multi-disciplinaire :

- comprendre sa contribution spécifique dans le projet ;
- transmettre de manière claire son apport (outils de formalisation) ;
- assurer les coordinations nécessaires.

**Objectifs d'apprentissage de la spécialité en termes de compétences métier**

- Technologies et méthodes de conception et développement : bases de données relationnelles, normes et outils pour documents structurés, conception de produits informationnels ;
- Connaissances des produits et outils industriels en gestion d'information et en traitement des documents.
- Capacité de maîtriser la gestion de projets
- Traitement du document numérique

## Débouchés

**Métiers auxquels la spécialité permet d'accéder directement**

Ingénieur linguiste, terminologue, gestionnaire de site web multilingue, lexicologue, chef de projet multimédia, traducteur

| Code | Intitulé |
|---|---|
| 32213 | Webmaster |
| 32214 | Documentaliste spécialisé(e) (dans un domaine)<br>ou Responsable du service documentation |
| 32241 | Traducteur |
| 32321 | Ingénieur de la connaissance |
| 32331 | Chef de projet Internet ou multimédia |
| 32341 | Architecte système d'information ou Responsable d'études informatiques |
| 35152 | Lexicologie, terminologie |

**Relation avec les milieux professionnels** :

EDF, Mondeca, Temis, Arisem, XEROX, PUF, Larousse, Le Robert, Performix, Syllabs, Quensis, Logosapience, Exalead, Thales, France Telecom, Limsi, Bowne Global Solutions, Softissimo, TRADOS, SDL, LIP6, ATILF

## Organisation globale des enseignements du master

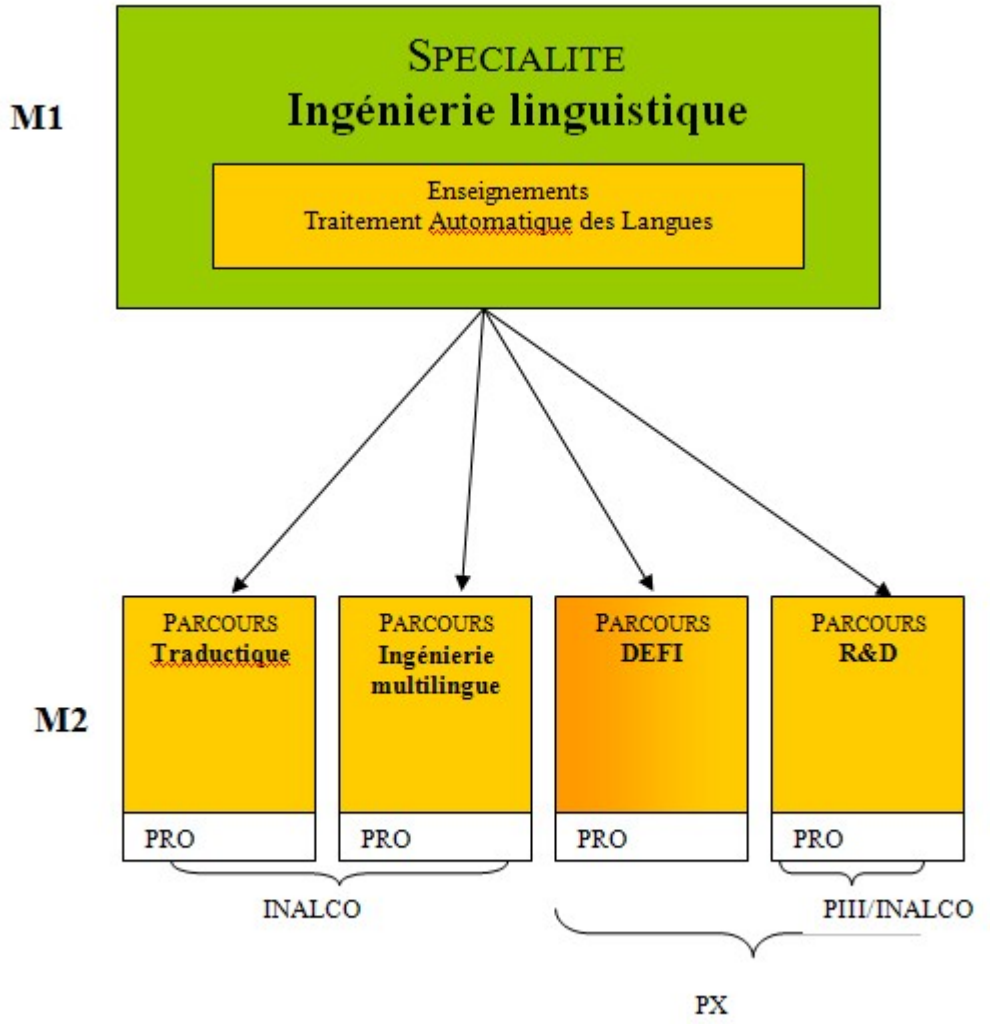
La spécialité s'organise selon quatre parcours, adossés chacun à un (ou plusieurs établissements universitaires) :

- le parcours *DEFI* (Documents Electroniques et Flux d'Informations), basé à Paris Ouest Nanterre La Défense, qui prend la suite de la spécialité *DEFI* du master mention Sciences du langage (URL du parcours *DEFI* : http://www.u-paris10.fr/defi/);

- le parcours *Ingénierie multilingue*, basé à l'INALCO ;

- le parcours *Traductique*, également basé à l'INALCO ;

- le parcours *Recherche et Développement*, basé dans les trois établissements.



La première année, les UE sont identiques entre les quatre parcours (*cf*. tableau ci-dessous). Les EC qui composent les UE sont soit des EC communes, proposées à l'ensemble des étudiants, soit des EC au choix qui peuvent être pris entre l'un ou l'autre des établissements.

A noter que les UE « Document numérique et Informatique » 1 et 2 sont également proposées dans la spécialité DEFI du master mention Sciences de l'information et de la communication.

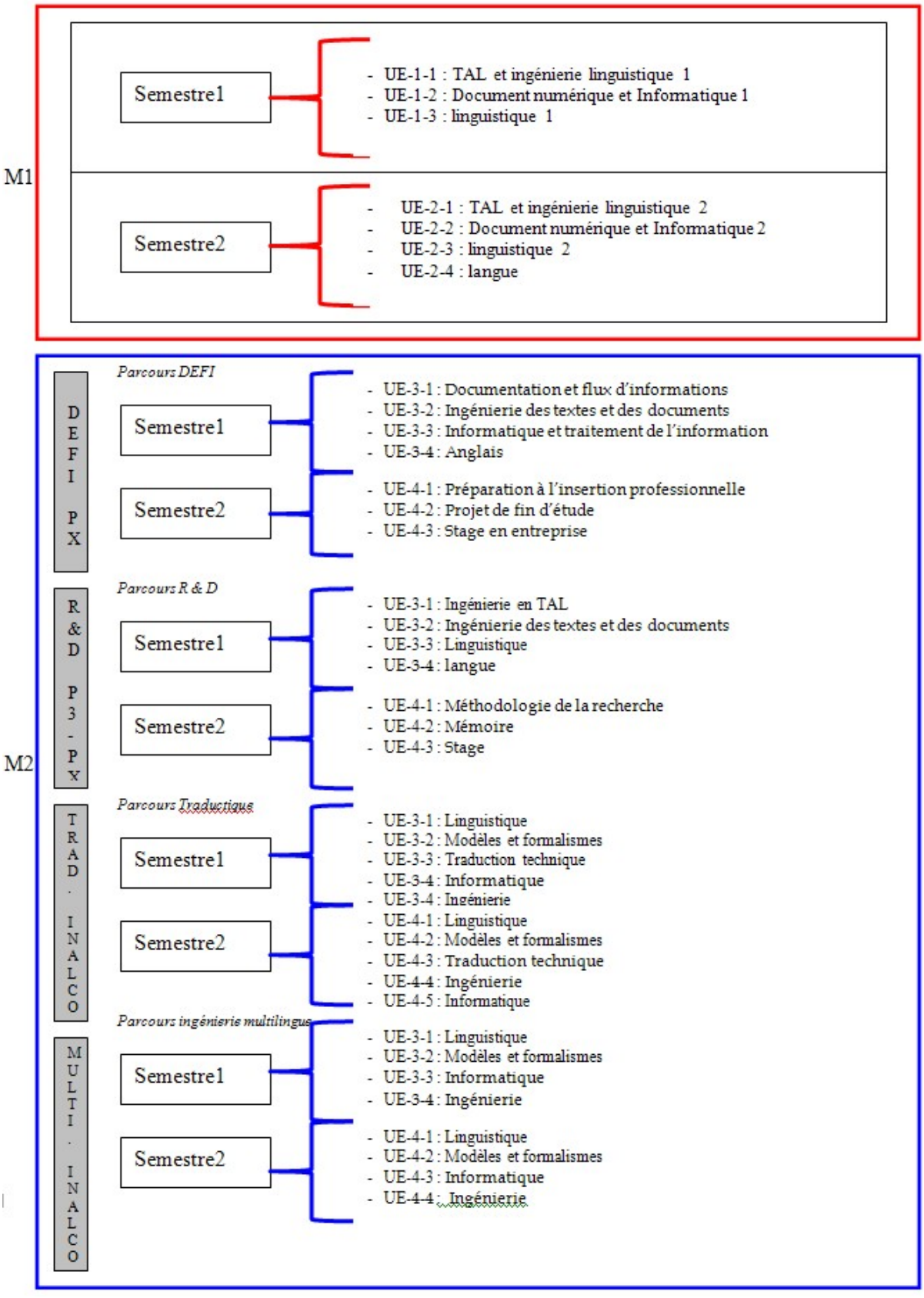
En deuxième année, les parcours se diversifient nettement.

Le parcours *DEFI* et le parcours *Recherche et développement* ont en commun au premier semestre l'UE « Ingénierie des textes et des documents ». Ils se distinguent par les UE : « Documentation et flux d'informations » et « Informatique et traitement de l'information » en DEFI, « Ingénierie linguistique et TAL » et « Linguistique » en Recherche et développement. Le deuxième semestre est entièrement différencié, puisque dans le parcours *Recherche et développement* il y a obligatoirement un mémoire, et que l'UE « Préparation à l'insertion professionnelle » est remplacée par de la méthodologie de la recherche.

La deuxième année du parcours *DEFI* est par ailleurs identique à la deuxième année de la spécialité *DEFI* de la mention Sciences de l'information et de la communication.

Enfin, les parcours *Ingénierie multilingue* et *Traductique* se distinguent en deuxième année par des enseignements de « Techniques de traduction » pour ce dernier parcours, ce qui a pour répercussion que les UE « Linguistique », « Modèles et formalismes », « Informatique » et « Ingénierie » ont un moindre poids dans ce parcours.

## Tableau synthétique des différents parcours en M1 et M2



# MASTER 1ère année

**4 parcours :**
*PRO* parcours DEFI (noté **D**)
*PRO* parcours RECHERCHE et DEVELOPPEMENT en TAL (noté **R**)
*PRO* parcours TRADUCTIQUE (noté **T**)
*PRO* parcours INGENIERIE MULTILINGUE (noté **I**)

## Semestre 1

| ECTS | Parcours D | Parcours R | Parcours I | Parcours T | ECTS |
|---|---|---|---|---|---|
| 15 | TAL ET INGIENIERIE LINGUISTIQUE 1 | | | | 15 |
| 3<br>3<br>3 | gestion informatique du multilinguisme<br>informatique et phonétique<br>programmation et projet encadré 1 | | | | |
| 3 | analyse syntaxique automatique | | principes des bases de données 1 | | |
| 3 | grammaire formelle | | langages réguliers et langages hors contexte1 | | |
| 6 | DOCUMENT NUMERIQUE ET INFORMATIQUE 1 | | | | 6 |
| 3 | bases de données | | | | |
| 3 | programmation et algorithmique 1 | | représentation et algorithmes<br>programmation et abstraction des données 1<br>remise à niveau info | | |
| 9 | LINGUISTIQUE 1 | | | | 9 |
| 3<br>3<br>3 | 3 ECTS de linguistique<br>génération automatique<br>syntaxe formelle (Paris X) | | Compétence avancée en langue<br>lexique et morphologie<br>pratiques textuelles et traduction | | |

**4 parcours :**
*PRO* parcours DEFI (noté **D**)
*PRO* parcours RECHERCHE et DEVELOPPEMENT en TAL (noté **R**)
*PRO* parcours TRADUCTIQUE (noté **T**)
*PRO* parcours INGENIERIE MULTILINGUE (noté **I**)

## Semestre 2

| ECTS | Parcours D | Parcours R | Parcours I | Parcours T | ECTS |
|---|---|---|---|---|---|
| 12 | TAL ET INGIENIERIE LINGUISTIQUE 2 | | | | 12 |
| 3 | statistique et analyse multidimensionnelle | | | | |
| 3 | programmation et projet encadré 2 | | | | |
| 3 | outils de traitements de corpus | | | | |
| 3 | corpus parallèles et comparables | | | | |
| | | | langages réguliers et langages hors contexte2 | | |
| 9 | DOCUMENT NUMERIQUE ET INFORMATIQUE 2 | | | | 9 |
| 3 | recherche et extraction d'information | | | | |
| 3 | document structuré | | | | |
| 3 | programmation et algorithmique 2 | | principes des bases de données 2<br>remise à niveau info<br>programmation logique<br>programmation et abstraction des données 2 | | |
| 6 | LINGUISTIQUE 2 | | | | 6 |
| 3 | syntaxe formelle (Paris III) | | sémantique textuelle | | |
| 3 | 3 ECTS de linguistique | | techniques de traduction | | |
| 3 | LANGUE | | | | 3 |
| 3 | langue | | compétence avancée en langue | | |

## Master 1, tous parcours Paris III, Paris Ouest Nanterre La Défense

| *Semestre 1* | | |
|---|---|---|
| **TAL et ingénierie linguistique 1** | **15** | |
| Grammaires formelles | 3 | Paris X |
| Analyse syntaxique automatique | 3 | Paris X |
| Informatique et phonétique | 3 | Paris III |
| Gestion informatique du multilinguisme | 3 | INALCO |
| Projet encadré 1 | 3 | Paris III/ INALCO |
| **Linguistique 1** | **9** | |
| Génération automatique de textes | 3 | Paris III |
| Syntaxe Formelle (Paris X) | 3 | Paris X |
| (+ 3 crédits de linguistique) | 3 | |
| **Document numérique et Informatique 1** | **6** | |
| Bases de données | 3 | Paris X |
| Programmation et algorithmique 1 | 3 | Paris III |
| *Semestre 2* | | |
| **TAL et ingénierie linguistique 2** | **12** | |
| Statistique et analyse multidimensionnelle | 3 | Paris III |
| Corpus parallèles et comparables | 3 | INALCO |
| Projet encadré 2 | 3 | Paris III/ INALCO |
| Outils de traitement de corpus | 3 | INALCO |
| **Linguistique 2** | **6** | |
| Syntaxe formelle (Paris III) | 3 | Paris III |
| (+ 3 crédits de linguistique) | 3 | |
| **Document numérique et Informatique 2** | **9** | |
| Document structuré | 3 | Paris III |
| Programmation et algorithmique 2 | 3 | Paris X |
| Recherche et extraction d'information | 3 | Paris X |
| **Langue** | **3** | P III/ PX/ INALCO |

## Master 1, INALCO

| *Semestre 1* | | |
|---|---|---|
| **TAL et ingénierie linguistique 1** | **15** | |
| Langages réguliers et langages hors contexte 1 | | INALCO |
| Informatique et phonétique | | Paris III |
| Gestion informatique du multilinguisme | | INALCO |
| Projet encadré 1 | | Paris III/ INALCO |
| Principes des bases de données | | INALCO |
| **Linguistique 1** | **9** | |
| Pratique textuelle | | INALCO |
| Lexique et morphologie | | INALCO |
| Compétence avancée en langue | | INALCO |
| **Document numérique et Informatique 1** | **6** | |
| Bases de données | | Paris X |
| Représentations et algorithmes | | INALCO |
| Programmation et abstraction des données 1 | | INALCO |
| Remise à niveau info | | INALCO |
| *Semestre 2* | | |
| **TAL et ingénierie linguistique 2** | **12** | |
| Statistique et analyse multidimensionnelle | | Paris III |
| Corpus parallèles et comparables | | INALCO |
| Projet encadré 2 | | Paris III/ INALCO |
| Outils de traitement de corpus | | INALCO |
| Langages réguliers et langages hors contextes 2 | | INALCO |
| **Linguistique 2** | **6** | |
| Pratique textuelle | | INALCO |
| Lexique et morphologie | | INALCO |
| **Document numérique et Informatique 2** | **9** | |
| Document structuré | | Paris III |
| Recherche et extraction d'information | | INALCO |
| Programmation logique | | INALCO |
| Programmation et abstraction des données 2 | | INALCO |
| Remise à niveau info | | INALCO |
| **Langue** | **3** | |
| Compétence avancée en langue | | INALCO |

## Code des enseignements du M1 (semestre 1)

| Code Paris X | Code Paris III | Code INALCO | Unité D'enseignement |
|---|---|---|---|
| **LMTUD110** | | | TAL ET INGIENIERIE LINGUISTIQUE 1 |
| LMLGAT15 | | | gestion informatique du multilinguisme |
| LMLGAT13 | | | informatique et phonétique |
| LMLGAT12 | | | analyse syntaxique automatique |
| LMLGAT14 | | | programmation et projet encadré 1 |
| LMLGAT11 | | | grammaire formelle |
| | | | principes des base de données |
| | | | langages réguliers et langages hors contexte 1 |
| **LMTUT130** | | | DOCUMENT NUMERIQUE ET INFORMATIQUE 1 |
| LMLGAT18 | | | bases de données |
| LMLGAT19 | | | programmation et algorithmique 1 |
| | | | représentations et algorithmes |
| | | | programmation et abstraction des données 1 |
| | | | remise à niveau info |
| **LMTUD120** | | | LINGUISTIQUE 1 |
| LMLGAT16 | | | génération automatique |
| | | | syntaxe formelle (Paris X) |
| | | | pratique textuelle |
| | | | lexique et morphologie |
| | | | compétence avancée en langue |

## Code des enseignements du M1 (semestre 2)

| Code Paris X | Code Paris III | Code INALCO | Unité D'enseignement |
|---|---|---|---|
| **LMTUD210** | | | TAL ET INGIENIERIE LINGUISTIQUE 2 |
| LMLGAT21 | | | statistique et analyse multidimensionnelle |
| LMLGAT24 | | | programmation et projet encadré 2 |
| LMLGAT23 | | | outils de traitements de corpus |
| LMLGAT22 | | | Corpus parallèles et comparables |
| | | | langages réguliers et langages hors contextes 2 |
| **LMTUD230** | | | DOCUMENT NUMERIQUE ET INFORMATIQUE 2 |
| LMLGAT28 | | | recherche et extraction d'information |
| LMLGAT30 | | | document structuré |
| LMLGAT29 | | | programmation et algorithmique 2 |
| | | | programmation logique |
| | | | programmation et abstraction |
| | | | remise à niveau info |
| **LMTUD220** | | | LINGUISTIQUE 2 |
| LMLGAT26 | | | syntaxe formelle (Paris 3) |
| | | | sémantique textuelle |
| | | | techniques de traduction |
| **LMTUT240** | | | LANGUE |
| | | | compétence langue |
| LMLGAT40 | | | langue |

## MASTER 2ème année

4 blocs distincts

| | | |
|---|---|---|
| *pro* | **Parcours D** | le M2 Parcours *DEFI* Paris X |
| *pro* | **Parcours R** | le M2 Parcours *Recherche & Développement* Paris III/Paris X/INALCO |
| *pro* | **Parcours T** | le M2 Parcours *Traductique* INALCO |
| *pro* | **Parcours I** | le M2 Parcours *Ingénierie multilingue* INALCO |

Les contenus de ces 4 parcours en M2 sont décris *infra*.

## Parcours D : M2 DEFI, Paris Ouest Nanterre La Défense

### Semestre 1 (30 ECTS)

| Code Paris X | *Unités d'enseignement* | *crédits* |
|---|---|---|
| | **DOCUMENTATION ET FLUX D'INFORMATION** | ***6*** |
| | Structuration et nature de l'information | 3 |
| | Veille et intelligence économique | 3 |
| | **INGENIERIE DES TEXTES ET DES DOCUMENTS** | ***9*** |
| | Traitement automatique de la langue et Ingénierie linguistique | 3 |
| | Méthodes du traitement automatique des langues | 3 |
| | Document structuré : XML | 3 |
| | **INFORMATIQUE ET TRAITEMENT DE L'INFORMATION** | ***12*** |
| | Programmation et programmation orientée objet | 4 |
| | Base de données et Web dynamique | 4 |
| | Gestion de l'information électronique | 4 |
| | **ANGLAIS** | ***3*** |
| | Anglais | 3 |

### Semestre 2 (30 ECTS)

| Code Paris X | *Unités d'enseignement* | *crédits* |
|---|---|---|
| | **PREPARATION A L'INSERTION PROFESSIONNELLES** | ***6*** |
| | Industrialisation des processus | 2 |
| | Gestion de projets | 2 |
| | Conférences professionnelles | **2** |
| | **PROJET DE FIN D'ETUDE** | ***9*** |
| | Projet de fin d'étude | 9 |
| | **STAGE EN ENTREPRISE** | ***15*** |
| | Stage en entreprise | 15 |

## Parcours R : M2 R&D, Paris Ouest Nanterre La Défense, Paris III, Inalco

### Semestre 1 (30 ECTS)

| Code P3 | Code Inalco | Code Paris X | *Unités d'enseignement* | *crédits* |
|---|---|---|---|---|
| | | **LMTUD311** | **INGENIERIE EN TAL** | ***9*** |
| | | | **3 enseignements** à prendre parmi ceux qui suivent | |
| | | | **SITE Paris Ouest Nanterre La Défense** | |
| | | LMLGAT31 | Modélisation des langues : mots et grammaire | 3 |
| | | LMLGAT35 | Formalismes syntaxiques et traitement automatique | 3 |
| | | | **SITE Paris III** | |
| | | LMLGAT32 | Analyse du discours et lexicométrie | 3 |
| | | | **SITE INALCO** | |
| | | LMLGAT33 | Écritures et multilinguisme | 3 |
| | | | Multimodalité du document numérique | 3 |
| | | LMLGAT34 | Sémantique des textes multilingues | 3 |
| | | **LMTUD321** | **INGENIERIE DES TEXTES ET DES DOCUMENTS** | ***9*** |
| | | LMLGAT36 | Traitement automatique de la langue et Ingénierie linguistique | 3 |
| | | LMLGAT37 | Méthodes du traitement automatique des langues | 3 |
| | | LMDEF | Document structuré : XML | 3 |
| | | **LMTUD331** | **LINGUISTIQUE** | ***9*** |
| | | | **2 ou 3 enseignements** de linguistique<br>à prendre en **accord avec le directeur de recherche** | 9 |
| | | **LMTUT341** | **LANGUE** | |
| | | | Langue vivante | 3 |

### Semestre 2 (30 ECTS)

| Code P3 | Code Inalco | Code Paris X | *Unités d'enseignement* | *crédits* |
|---|---|---|---|---|
| | | **LMTUD421** | **STAGE** | ***9*** |
| | | | Stage en laboratoire ou en entreprise | 9 |
| | | **LMTUT431** | **MEMOIRE** | ***15*** |
| | | | Mémoire de recherche | 15 |
| | | **LMTUD411** | **METHODOLOGIE DE LA RECHERCHE** | ***6*** |
| | | | Méthodologie de la recherche | 6 |

## Parcours T : M2 Traductique, Inalco

### Semestre 1 (30 ECTS)

| Code INALCO | *Unités d'enseignement* | *crédits* |
|---|---|---|
| | **LINGUISTIQUE** | ***6*** |
| LTE09100<br>LTP09100<br>LTP09101 | Sémantique des textes multilingues 1<br>Lexicologie, terminologie, dictionnairique 1<br>Genres, textes et usages 1 | |
| | **MODELES ET FORMALISMES** | ***6*** |
| LTP09103<br>LTE09103 | Acquisition, modélisation et représentation des connaissances<br>Documents structurés | |
| | **TRADUCTION TECHNIQUE** | ***9*** |
| LTP09109<br>LTP09110<br>LTP09111 | Traduction technique 1<br>Traductologie 1<br>Conduite de projet de traduction 1 | |
| | **INFORMATIQUE** | ***3*** |
| LTP09108<br>LTP09105 | Programmation shell<br>Bases de données pour le web | |
| | **INGENIERIE** | ***6*** |
| LTE09101<br>LTP09197<br>LTR09103<br>LTP09112 | Multimodalité du document numérique<br>Outils de TAO 1<br>Ecritures et multilinguisme<br>Traitement statistique de corpus | |

### Semestre 2 (30 ECTS)

| Code INALCO | *Unités d'enseignement* | *crédits* |
|---|---|---|
| | **LINGUISTIQUE** | ***3*** |
| LTE10100<br>LTP10100 | Sémantique des textes multilingues 2<br>Lexicologie, terminologie, dictionnairique 2 | |
| | **MODELES ET FORMALISMES** | ***3*** |
| LTP10102 | Indexation et gestion électronique de documents | |
| | **TRADUCTION TECHNIQUE** | ***6*** |
| LTP10109<br>LTP10110<br>LTP10111 | Traduction technique 2<br>Traductologie 2<br>Conduite de projet de traduction 2 | |
| | **INGENIERIE** | ***3*** |
| LTP10106<br>LTP10107 | Techniques web<br>Outils de TAO 2 | |
| | **STAGE + MEMOIRE** | ***15*** |
| | STAGE + MEMOIRE | |

## Parcours I : M2 Ingénierie Multilingue, Inalco

### Semestre 1 (30 ECTS)

| Code INALCO | *Unités d'enseignement* | *crédits* |
|---|---|---|
| | **LINGUISTIQUE** | ***6*** |
| LTE09100 | Sémantique des textes multilingues 1 | |
| LTP09100 | Lexicologie, terminologie, dictionnairique 1 | |
| LTP09101 | Genres, textes et usages 1 | |
| | **MODELES ET FORMALISMES** | ***9*** |
| LTP09102 | Calculabilité | |
| LTP09103 | Acquisition, modélisation et représentation des connaissances | |
| LTE09103 | Documents structurés | |
| | **INFORMATIQUE** | ***6*** |
| LTP09104 | Programmation objet 1 | |
| LTP09105 | Langages de scripts | |
| | **INGENIERIE** | ***9*** |
| LTE09101 | Multimodalité du document numérique | |
| LTP09107 | Outils de TAO 1 | |
| | Analyse du discours et lexicométrie | |
| LTR09103 | Ecritures et multilinguisme | |
| | Outils de traitement de corpus | |
| LTP09112 | Traitement statistique de corpus | |

### Semestre 2 (30 ECTS)

| Code INALCO | *Unités d'enseignement* | *crédits* |
|---|---|---|
| | **LINGUISTIQUE** | ***3*** |
| LTE10100 | Sémantique des textes multilingues 2 | |
| LTP10100 | Lexicologie, terminologie, dictionnairique 2 | |
| | **MODELES ET FORMALISMES** | ***3*** |
| | Analyse robuste | |
| | **INFORMATIQUE** | ***6*** |
| LTP10104 | Programmation objet 2 | |
| LTP10103 | Programmation itérative et récursive | |
| LTP10105 | Bases de données sur le web | |
| | **INGENIERIE** | ***3*** |
| LTP10106 | Techniques web | |
| LTP10107 | Outils de TAO 2 | |
| | **STAGE + MEMOIRE** | ***15*** |
| | STAGE + MEMOIRE | |

## Contenu des unités d'enseignement

**UE de M1 S1 :**
*TAL et ingénierie linguistique 1* Cette UE donne un aperçu d'un certain nombre d'outils et de méthodes du traitement automatique des langues. Sont étudiées les Grammaires formelles (ou les langages réguliers), l'analyse syntaxique automatique, Informatique et phonétique, la gestion informatique du multilinguisme. La réalisation d'un projet est amorcée.

*Linguistique 1* Cette UE a pour objectif d'accroître les connaissances en linguistique générale des étudiants. En dehors d'un enseignement sur la génération automatique de textes, ils pourront prendre à leur choix 6 crédits de linguistique.

*Document numérique et Informatique 1* Il s'agit, outre la familiarisation avec la notion de bases de données, le début d'un apprentissage systématique et raisonné de la programmation.

**UE de M1 S2 :**
*TAL et ingénierie linguistique 2* On continue d'appréhender les outils et méthodes du traitement automatique des langues, notamment les méthodes statistique, le traitement de corpus, la syntaxe formelle (ou des compléments sur les langages réguliers). Le projet commencé au semestre 1 est achevé.

*Linguistique 2* Les étudiants poursuivent le perfectionnement de leur culture en linguistique générale, en suivant un enseignement de syntaxe formelle à quoi s'ajoutent trois crédits à prendre au choix.

*Document numérique et Informatique 2* Outre la poursuite de l'apprentissage de la programmation, les étudiants se familiarisent avec les techniques de Recherche et extraction d'information et de traitement des documents structurés.

**UE de M2 DEFI :**
*Documentation et flux d'informations* On étudie les fondements sur la structuration et la nature de l'information, et il est développé ce que sont la veille et l'intelligence économique

*Informatique et traitement de l'information* Les connaissances en programmation sont prolongées par un apprentissage de la programmation orientée objet, par la manipulation des chaînes de traitement sous Unix. Les connaissances en bases de données sont prolongées par la connaissance de la mise en place de sites Web dynamiques. Enfin toute la chaîne de la gestion de l'information électronique est présentée.

*Préparation à l'insertion professionnelle* On explique comment sont industrialisés les processus, notamment en ingénierie linguistique. De même est présentée la gestion des projets, et des conférences sont données par des professionnels.

*Projet de fin d'études* Par groupes de 4 à 6 étudiants est réalisé un projet qui nécessite une étude de la tâche à accomplir, la définition d'un cahier des charges, la répartition du travail, et qui fait appel à un certain nombre des connaissances techniques acquises dans les autres UE du master.

*Stage en entreprise* L'étudiant s'intègre réellement dans une entreprise, et il présente son expérience dans un rapport de stage soutenu oralement.

**UE de M2 DEFI et R&D :**
*Ingénierie des textes et des documents* On étudie et manipule un certain nombre de logiciels de traitement automatique de la langue et notamment des outils statistiques. On se perfectionne dans le traitement des expressions régulières, et on va aussi loin que possible dans l'appréhension des outils de traitement des documents structurés.

**UE de M2 R&D :**
*Ingénierie en TAL* Il s'agit, pour les étudiants, de se familiariser avec la démarche de recherche en traitement automatique des langues et/ou en linguistique formelle, à travers l'étude systématique de quelques problèmes bien définis et d'approches bien identifiées.

*Linguistique* L'objectif est d'approfondir la culture en linguistique des étudiants, en les mettant au fait d'un certain nombre de recherches actuelles dans le domaine.

*Méthodologie de la recherche* L'étudiant, à travers notamment une réflexion sur son propre travail d'élaboration d'un mémoire et celui des autres étudiants, progresse dans ses capacités à faire de la recherche dans le domaine.

*Stage en laboratoire ou en entreprise* Étude d'un problème dans un contexte de coopération.

*Mémoire* Rédaction d'un mémoire de recherche soutenu oralement devant un jury.

**UE de M2 Ingénierie multilingue (INALCO):**

| | |
|---|---|
| Linguistique (I) | Sémantique des textes multilingues |
| | Lexicologie, terminologie, dictionnairique |
| | Genres, textes et usages |
| Modèles et formalismes (I) | Documents structurés |
| | Calculabilité |
| | Acquisition et modélisation des connaissances |
| Informatique (I) | programmation objet 1 |
| | Langages de script |
| Ingénierie (I) | Multimodalité du document numérique 1 |
| | Analyse du discours et lexicométrie |
| | Outils de TAO 1 |
| | Outils de traitement de corpus |
| | Traitement statistique de corpus |
| | Ecritures et multilinguisme |
| Linguistique (I) | Sémantique des textes multilingues 2 |
| | Lexicologie, terminologie, dictionnairique 2 |
| Modèles et formalismes (I) | Indexation et gestion électronique de documents |
| Informatique (I) | Analyse robuste |
| | Programmation itérative et récursive |
| | Programmation objet 2 |

| | |
|---|---|
| | Bases de données sur le web |
| Ingénierie (I) | techniques web |
| | Outils de TAO 2 |

**UE de M2 Traductique (INALCO):**

| | |
|---|---|
| Linguistique (T) | Sémantique des textes multilingues 1 |
| | Lexicologie, terminologie, dictionnairique 1 |
| | Genres, textes et usages |
| Modèles et formalismes (T) | document structurés |
| | Acquisition et modélisation des connaissances |
| Traduction technique (T) | Traductologie 1 |
| | Conduite de projets de traduction 1 |
| Informatique (T) | Bases de données pour le web |
| | programmation shell |
| Ingénierie (T) | Multimodalité du document numérique |
| | Outils de TAO 1 |
| | Traitement statistique de corpus |
| | Ecritures et multilinguisme |
| Linguistique (T) | Sémantique des textes multilingues 2 |
| | Lexicologie, terminologie, dictionnairique 2 |
| Modèles et formalismes (T) | Indexation et gestion électronique de documents |
| Traduction technique (T) | Traduction technique 2 |
| | Traductologie 2 |
| | Conduite de projets de traduction 2 |
| Ingénierie (T) | techniques web |
| | outils de TAO 2 |

# Planning des cours du Tronc commun du Master Ingénierie Linguistique

## MASTER 1

INALCO PARISX PARIS 3

### Semestre 1

| | 8h | 9h | 10h | 11h | 12h | 13h | 14h | 15h | 16h | 17h | 18h |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Lundi | | | | | | | | | | | |
| Mardi | | | | | | | Gestion Info. du multilinguisme | | | | |
| Mercredi | | Projet encadré | | | Traitement de l'Oral | | Génération | | | Algo/Programmation | |
| Jeudi | | | | | | | | | | | |
| Vendredi | | Syntaxe formelle | | Analyse Syntax. Auto. | | BDD linguistes | | Gram. Formelles | | | |

### Semestre 2

| | 8h | 9h | 10h | 11h | 12h | 13h | 14h | 15h | 16h | 17h | 18h |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Lundi | | Outils Corpus. Corpus Parallèles comparables | | | | | | | | | |
| Mardi | | | | | | | | | | | |
| Mercredi | | Doc. Struct. P3 | | Projet encadré | | | Lexicométrie/ Stat. Textuelles | | | Synt. Form. P3 | |
| Jeudi | | | | | | | | | | | |
| Vendredi | | | Rech. d'information | | | Anglais Paris X | | Algo/Programmation | | | |

Ce planning n'intègre pas les enseignements optionnels

## MASTER 2

INALCO PARISX PARIS 3

### Semestre 1

| | 8h | 9h | 10h | 11h | 12h | 13h | 14h | 15h | 16h | 17h | 18h |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Lundi | | | | | | TAL et Ingénierie linguistique | | | Formalismes et TA | | |
| Mardi | | | | | | | | | | | |
| Mercredi | | Document structuré : XML | | | | | | | | | |
| Jeudi | | | | | | | | Anal. du disc. et Lexico. | | | |
| Vendredi | | | | | | Modélisation des langues | | | | | |

### Semestre 2

| | 8h | 9h | 10h | 11h | 12h | 13h | 14h | 15h | 16h | 17h | 18h |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Lundi | | | | | | | | | | | |
| Mardi | | | | | | | | | | | |
| Mercredi | | | | | | | | | | | |
| Jeudi | | | | | | | | | | | |
| Vendredi | | | | | | | | | | | |

Ce planning n'intègre pas les enseignements optionnels

## Equipe pédagogique

Nom : CLAVERIE Prénom : Camille
Université / UFR de rattachement : LLPHI
Equipe de recherche : CRIS

Nom : CORI Prénom : Marcel
Université / UFR de rattachement : Université Paris 10 – Nanterre, UFR LLPhi
Equipe de recherche : MoDyCo

Nom : DESMETS Prénom : Marianne
Université / UFR de rattachement : Université Paris 10 – Nanterre, UFR LLPhi
Equipe de recherche : MoDyCo

Nom : DAUBE Prénom : Jean-Michel
Université / UFR de rattachement : INaLCO
Equipe de recherche : ER-TIM (EAD 2540)

Nom : FANTON Prénom : Michel
Université / UFR de rattachement : INaLCO
Equipe de recherche : ER-TIM (EAD 2540)

Nom : FLEURY Prénom : Serge
Université / UFR de rattachement : Université Paris III Sorbonne nouvelle
Equipe de recherche : EA2290, SYLED/CLA$^2$T

Nom : GENDROT Prénom : Cédric
Université / UFR de rattachement : Université Paris III Sorbonne nouvelle
Equipe de recherche : Laboratoire de Phonétique et de Phonologie, UMR7018

Nom : GERDES Prénom : Kim
Université / UFR de rattachement : Université Paris III Sorbonne nouvelle , ILPGA
Equipe de recherche : Laboratoire de Phonétique et de Phonologie, UMR7018

Nom : KAHANE Prénom : Sylvain
Université / UFR de rattachement : Université Paris 10 – Nanterre, UFR LLPhi
Equipe de recherche : MoDyCo

Nom : MINEL Prénom : Jean-Luc
Université / UFR de rattachement : Université Paris 10 – Nanterre
Equipe de recherche : MoDyCo

Nom : MOREAUX Prénom : Marie-Anne
Université / UFR de rattachement : INaLCO
Equipe de recherche : ER-TIM (EAD 2540)

Nom : RIGOT Prénom : HUGUETTE
Université / UFR de rattachement : INaLCO
Equipe de recherche : ER-TIM (EAD 2540)

Nom : SALEM Prénom : André
Université / UFR de rattachement : Université Paris III Sorbonne nouvelle, ILPGA
Equipe de recherche : EA2290, SYLED/CLA$^2$T

Nom : SAMVELIAN Prénom : Pollet
Université / UFR de rattachement : Université Paris III Sorbonne nouvelle
Equipe de recherche : Mondes iranien et indien - UMR 7528

Nom : SLODZIAN Prénom : Monique
Université / UFR de rattachement : INaLCO

Equipe de recherche : ER-TIM (EAD 2540)

Nom : ZWEIGENBAUM Prénom : Pierre
Université / UFR de rattachement : INaLCO
Equipe de recherche : ER-TIM (EAD 2540)

## Descriptif et horaires des cours (1ère et 2ème années), tronc commun

Les horaires et lieux des cours présentés ci-dessous seront disponibles au moment de la rentrée universitaire (ils seront mis en ligne sur le site pluriTAL et diffusés sur la liste pluriTAL). On obtiendra des renseignements précis et à jour concernant ces cours en s'adressant aux secrétariats des UFRs concernés.

### Descriptif et horaires des cours du master 1ère année

---

### Grammaires formelles

**Enseignant** : Sylvain Kahane (Paris Ouest Nanterre La Défense)
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : vendredi 15h00-17h00

Le cours présente la théorie des langages formels et les grammaires formelles de référence en linguistique : les incontournables grammaires de réécriture de Chomsky, les grammaires lexicalisées avec les grammaires catégorielles, les grammaires de dépendance et les TAG (Grammaire d'adjonction d'arbres), la synchronisation de grammaires avec LFG (Grammaire Lexicale Fonctionnelle). La nature des représentations des énoncés manipulées par ces grammaires sera également abordée : arbres, arbres ordonnés, graphes acycliques.

**Bibliographie**

Abeillé Anne, Les nouvelles syntaxes : grammaires d'unification et analyse du français, Armand Colin, 1993.
Kahane Sylvain, Grammaires de dépendance formelles et théorie Sens-Texte, Tutoriel, Actes de TALN 2001, vol. 2, Tours, 2001, 60 pages, www.kahane.fr.

---

### Analyse syntaxique automatique

**Enseignant** : Marcel Cori (Paris Ouest Nanterre La Défense)
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : vendredi 10h30-12h30

On indique quelles sont les principales composantes qui entrent en jeu dans l'analyse syntaxique automatique : formalisme, grammaire, lexique, traits. On expose les différents types de méthodes (descendantes ou ascendantes, avec retour en arrière, en parallèle ou déterministe). On oppose l'analyse syntaxique théorique à l'analyse robuste, qui opère sur de « vraies » productions langagières, et qui en général reconnaît des fragments de phrase plutôt que des phrases complètes.

---

### Gestion informatique du multilinguisme

**Enseignant** : Jean François Perrot (INALCO), Marie-Anne Moreaux (INALCO)
**Lieu** : INALCO (Nogent), salle 8 et salle info 2
**Horaire** : mardi 14h00-17h00

Ce cours sera centré sur le substrat informatique et webographique en cause dans les problèmes de représentation, codage et transmission de l'information multilingue.
L'objectif est de permettre l'acquisition et la pratique des connaissances nécessaires à l'échange réussi de documents numériques multilingues provenant de machines, plate-formes et formats différents.

## Traitement de l'oral / Informatique et phonétique

**Enseignant** : Cédric Gendrot (Paris III)
**Lieu** : ILPGA, LaboC
**Horaire** : mercredi 12h30-14h30

Ce cours vise à présenter quelques aspects oraux du TAL, à savoir la synthèse de la parole, la reconnaissance vocale, l'utilisation de grands corpus oraux transcrits, etc. La synthèse de la parole actuellement utilisée par les serveurs vocaux est abordée dans un premier temps ; les différentes étapes sont mises en pratique en TD. Cette optique phonétique permet de passer en revue différents langages inspirés de XML utilisés pour l'amélioration de la synthèse ou pour la mise en parallèle de contenus multimédia. Dans une deuxième partie, plusieurs aspects peuvent être envisagés (alternativement selon les années) : depuis l'obtention de grands corpus oraux transcrits (étiquetage lexical et morpho-syntaxique) à partir de la reconnaissance automatique de la parole jusqu'à l'analyse phonétique automatisée de données de parole.

## Programmation et projet encadré (semestre 1)

**Enseignant** : Jean Michel Daube/Rachid Belmouhoub (INALCO), Serge Fleury (Paris III)
**Lieu** : ILPGA, LaboC
**Horaire** : mercredi 08h30-11h30

Il s'agit d'apprendre à mettre en œuvre une chaîne de traitement textuel semi-automatique, depuis la récupération des données jusqu'à leur utilisation. Ce cours posera d'abord la question des objectifs linguistiques à atteindre (lexicologie, recherche d'information, traduction...) et fera appel aux méthodes et outils informatiques nécessaires à leur réalisation (récupération de corpus, normalisation des textes, segmentation, étiquetage, extraction, structuration et présentation des résultats...). Ce cours sera aussi l'occasion d'une évaluation critique des résultats obtenus, d'un point de vue quantitatif et qualitatif.
URL : http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/cours/masterproj.htm

## Bases de données

**Enseignant** : J.L Milnel (Paris Ouest Nanterre La Défense)
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : vendredi 13h00-15h00

La description d'une réalité langagière se fait souvent « à la main » : corpus saisi sous traitement de texte, observations faites également sous traitement de texte, etc. Néanmoins, ces descriptions non structurées sont difficilement analysables lorsque le volume de données devient important.
Les systèmes de bases de données relationnelles fournissent des outils logiciels adaptés aux traitements de grandes masses de données structurées. L'accent est mis sur la démarche de création et d'utilisation de bases de données pour des phénomènes langagiers, notamment par l'utilisation du langage SQL. Une partie du cours se déroule sur machine, avec utilisation d'un logiciel de gestion de bases de données. Le cours ne suppose néanmoins pas de connaissances informatiques préalables.

## Statistique et analyse multidimensionnelle

**Enseignant** : André Salem (Paris III)
**Lieu** : ILPGA, salle LaboC
**Horaire** : mercredi 14h00-17h00

Les approches quantitatives des corpus textuels sont présentées (historique et tendances récentes). Est abordée la question des unités pour la statistique textuelle (formes, lemmes, segments répétés, cooccurrences). La compréhension d'un certain nombre de propriétés statistiques des textes (Zipf-Pareto, courbes d'accroissement du vocabulaire) fournit le cadre d'analyse des constats effectués. Le cours introduit également aux méthodes de statistiques appliquées aux données textuelles : indices, distances, approches multidimensionnelles.

---

## Corpus parallèles et comparables // Outil de Traitement de Corpus

**Enseignant** : Pierre Zweigenbaum (INALCO)
**Lieu** : INALCO
**Horaire** : lundi 09h30-12h30

Ce cours vise à expliciter les objectifs sous-jacents à l'établissement de corpus parallèles (où des textes sont en rapport de traduction) et à exposer les techniques linguistiques et informatiques mises en œuvre pour réaliser un alignement à différents paliers du document (paragraphe, phrase, mot). A partir des limites des corpus parallèles, on expliquera le recours aux corpus comparables (traitant du même domaine et relevant des mêmes genres), et les outils de traitement associés.

Horaires des cours "Outils de traitement de corpus" puis "Corpus parallèles et comparables" : lundis 9.30 - 12.30, la Belle Gabrielle. Début des cours le 1er février
6 cours de 3 heures "Outils de traitement de corpus" : 1, 8, 15 février, 1, 8, 15 mars
6 cours de 3 heures "Corpus parallèles et comparables" : 22, 29 mars, 12 avril, 3, 10, 17 mai

---

## Recherche et extraction d'information

**Enseignant** : Florence Amardeilj
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : vendredi 10h00-12h00, salle 115

La recherche d'information est à la base des moteurs de recherche actuels. Elle recouvre les mécanismes et les modèles permettant de rapatrier les documents correspondant à une requête. L'extraction d'information produit des réponses plus petites qu'un document, répondant de manière plus précise aux demandes d'un utilisateur. Elle est à la base par exemple des systèmes retournant des réponses à des questions factuelles (« En quelle année la France a-t-elle gagné la Coupe du Monde de Football ? »)

---

## Syntaxes formelles

**Enseignant** : Marcel Cori (Paris Ouest Nanterre La Défense)
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : (semestre 1) vendredi 08h30-10h30

---

## Syntaxes formelles

**Enseignant** : Pollet SAMVELIAN (Paris III)
**Lieu** : ILPGA
**Horaire** : mercredi 17h00-18h30 (semestre 2)

Ce séminaire se donne un double objectif :

1. Familiariser les étudiants avec un formalisme grammatical particulier, en l'occurrence le modèle HPSG (Head-Driven Phrase Structure Grammar, tr. fr.: grammaires syntagmatiques guidées par les têtes).

2. Mener une réflexion générale, par le biais de la comparaison entre le modèle présenté et l'une des variantes du modèle génératif-transformationnel (P&P), sur les différentes possibilités d'analyses pour un même phénomène syntaxique, selon les caractéristiques du modèle que l'on adopte.

HPSG appartient au paradigme des grammaires dites d'unification. Ces dernières, nées dans les années quatre-vingts, constituent une alternative au modèle génératif-transformationnel dans ses diverses variantes. Malgré leurs différences plus ou moins importantes, elles partagent généralement les caractéristiques suivantes, qui en font des outils adaptés pour l'implémentation en TAL :

- Absence de tout mécanisme transformationnel (donc de mouvement).
- Réhabilitation des descriptions linguistiques proches de la structure dite de surface.
- Recours au mécanisme d'unification de traits.

Les premières séances seront consacrées à la mise en place des notions communes à la plupart de ces modèles. On passera ensuite au modèle HPSG en abordant le(s) traitement(s) proposé(s) dans ce cadre pour les phénomènes suivants : montée et contrôle, dépendances à distance et extraction, interface morphologie-syntaxe (traitement des clitiques).
Pour le plan détaillé, l'organisation des séances et la bibliographie commentée se référer à :
http://www.cavi.univ-paris3.fr/ilpga/pages_personnelles/samvelian/index.htm

---

## Document structuré (M1)

**Enseignant** : Serge Fleury (Paris III)
**Lieu** : ILPGA, salle LaboC
**Horaire** : mercredi 09h00-10h30

Les textes sont des documents structurés. Un article comporte un titre, un ou des auteur(s), des sections, une bibliographie. La présentation permet d'appréhender cette structure (taille des caractères, jeu sur le gras, etc.). Lorsqu'on rend explicite cette structure (par le moyen de balisages en XML), on peut manipuler le texte comme unité structurée (extraire automatiquement les titres pour une table des matières, chercher les paragraphes introductifs, etc.). Le cours présente la manière de rendre explicite et fiable la structure des documents (en les assortissant d'une « grammaire textuelle » dite DTD). Il aborde les transformations réglées de textes qui deviennent possibles.

**Bibliographie**

P. Bonhomme, « Codage et normalisation de ressources textuelles », *in* Ingénierie des langues, J.-M. Pierrel (ed), p. 173-192, Hermès Science, 2000, Paris.

*Ressources fournie*s
Polycopié et outils sur pages WEB du cours :
URL : http://agora.univ-paris3.fr/course/view.php?id=1879

---

## Programmation et projet encadré (semestre 2)

**Enseignant** : Jean Michel Daube/Jean-Michel Daube (INALCO), Serge Fleury (Paris III)
**Lieu** : ILPGA, salle LaboC
**Horaire** : mercredi 10h30-13h30
*Cf* descriptif du premier semestre.
URL : http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/cours/masterproj.htm

---

## Programmation et algorithmique 1 et 2

**Enseignant** : Marcel Cori (Paris Ouest Nanterre La Défense), Kim Gerdes (Paris III)
**Lieu** : ILPGA
**Horaire** : au semestre 1 : mercredi 17h00-19h00 ILPGA, LaboC
au semestre 2 : vendredi 15h00-17h00, Paris X, salle 115

*Programmation et algorithmique 1* (Paris III, Kim Gerdes)
Ce cours constitue une introduction aux notions de programmation, théoriques et appliquées, adaptée aux besoins du linguiste informaticien. On abordera les idées sous-jacentes à l'algorithmique : la traduction d'un problème en instructions, la modularisation, les décisions, les boucles, représentées dans des organigrammes. Les exemples seront présentés en Python et étudiés sur machine. On développera des simples outils d'accès à des corpus, nécessaires pour des corrections et des comptages, auxquels tout linguiste est confronté régulièrement. On terminera sur les notions de hiérarchie de classes dans la programmation objet qu'on illustrera avec un travail sur les interfaces graphiques.

*Programmation et algorithmique 2* (Paris Ouest Nanterre La Défense, Marcel Cori)
Ce cours fait suite à *Programmation et algorithmique 1*. On approfondira les notions et méthodes vues au premier semestre en manipulant des structures de données dédiées à la linguistique et au traitement automatique des langues : arbres, automates, graphes, structures de traits. Des algorithmes et des programmes seront écrits sur ces structures, ce qui permettra la réalisation d'application élémentaires et plus élaborées de traitement automatique des langues (analyse morphologique, analyse syntaxique, génération automatique). Les programmes seront encore écrits dans le langage de programmation Python.

---

## Génération automatique de Textes

**Enseignant** : Kim Gerdes (Paris III)
**Lieu** : ILPGA, LaboC
**Horaire** : mercredi 14h30-16h00

La plupart des approches linguistiques partent de l'analyse de texte écrit (ou du son) pour en extraire une représentation morphologique, syntaxique, sémantique ou conceptuelle ; la correspondance saussurienne signifiant-signifié est habituellement décrite dans ce sens de l'interprétation. Nous verrons qu'une approche « générative », l'idée que la langue est en premier lieu la production langagière, nous ouvre de nouvelles perspectives linguistiques : Où commence la linguistique dans l'ensemble de processus cognitif mis en jeu ? Existe-t-il une représentation cognitive indépendante de la langue ? Qu'est-ce une structure sémantique ? Quoi dire ? Comment exprimer un sens avec des mots ? Comment structurer les idées en phrases ? Comment créer un discours cohérent ? Comment choisir les mots et les constructions syntaxiques ? Comment ordonner les mots ? Comment les grouper pour bien les prononcer ?
Ce cours aura à la fois une dimension théorique, sur ce que doit être un modèle d'une langue (nous présenterons en particulier la Théorie Sens-Texte), et une dimension applicative, en présentant des systèmes de génération de texte à partir de données brutes (météorologiques par exemple) ainsi que dans la traduction automatique.

**Bibliographie**

I. Mel'cuk (1997) Vers une linguistique Sens-Texte. Leçon inaugurale. Paris: Collège de France, 78 pages.
http://www.olst.umontreal.ca/FrEng/melcukColldeFr.pdf
E. Reiter & R. Dale (2000). Building natural language génération systems. Cambridge University Press, Cambridge.
D.D. McDonald (2000). « Natural Language Generation ». Chap. 7 : R. Dale, H. Moisl and H. Somers (éds.), Handbook of Natural Language
Processing, New York, Dekker.
L. Danlos, L. Roussarie (2000). « Génération Automatique de Textes » Chap. 14 : J.-M. Pierrel (éd), Ingénierie des langues, Paris, Hermès
Science.
W. Wahlster (2000). Verbmobil: foundations of speech-to-speech translation, Springer
M. Stede (1999). Lexical Semantics and Knowledge Representation in Multilingual Text Generation. Kluwer, Dordrecht.

---

## Descriptif et horaires des cours du master 2ème année

### Formalismes syntaxiques et traitement automatique

**Enseignant** : M. Cori, M. Desmets (Paris Ouest Nanterre La Défense)
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : lundi 16h00-18h00

---

### TAL et Ingénierie linguistique

**Enseignant** : J.L Minel (Paris Ouest Nanterre La Défense)
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : lundi 13h00-16h00

---

### Document structuré : XML (M2)

**Enseignant** : Jean-Luc Minel (Paris Ouest Nanterre La Défense)
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : mercredi 09h30-12h30

L'utilisation du langage XML pour décrire des documents semi-structurés nécessite généralement d'effectuer des transformations sur ces documents. L'utilisation des langage XSLT et XPATH est détaillée afin de montrer le type de transformation qu'il est possible de réaliser. Les langages RDF, RDFS et OWL (Web sémantique) qui appartiennent dorénavant à la galaxie XML sont présentés afin d'en montrer les principales finalités.

---

### Analyse du discours et lexicométrie

**Enseignant** : André Salem (Paris III)
**Lieu** : ILPGA, salle Benveniste
**Horaire** : jeudi 15h00-17h00

---

### Écritures et multilinguisme

**Enseignant** : M. Bottin (INALCO)
**Lieu** : INALCO
**Horaire** : (ce cours aura lieu au semestre 2)

Ce cours a pour objectif de sensibiliser les étudiants à la diversité des langues et de leurs écritures, en vue de ne pas plaquer les méthodes et outils informatiques d'une aire linguistique (monde anglo-saxon en particulier) sur une autre. On abordera en particulier les types d'écriture, les directions d'écriture, les problèmes de ligatures, majuscules, diacritiques, ponctuation, segmentation ... pour bien prendre la mesure des problèmes liés au multilinguisme.

---

## Multimodalité du document numérique

**Enseignant** : P. Stockinder, D. Piotrowski (INALCO)
**Lieu** : INALCO
**Horaire** : vendredi 10h00-12h00

Le document numérique n'est plus que faiblement textuel. Il associe également des images, du son. Il repose sur des interfaces d'accès qui influent sur sa compréhension et son utilisation. Le cours introduit à la dimension sémiotique de ces fonctionnements ainsi qu'à leur support technique.

---

## Sémantique des textes multilingues

**Enseignant** : M. Slodzian, E. Bourion, M. Valette, C. Podat (INALCO)
**Lieu** : INALCO
**Horaire** : jeudi 11h30-13h00

Le passage d'une langue à l'autre est filtré par les usages et les cultures associées à chacune des langues. De la structuration globale des textes aux séquences renvoyant à des entités du monde (institutions, événements), la sémantique doit rendre compte de cette dimension.

---

## Modélisation des langues : mots et grammaire

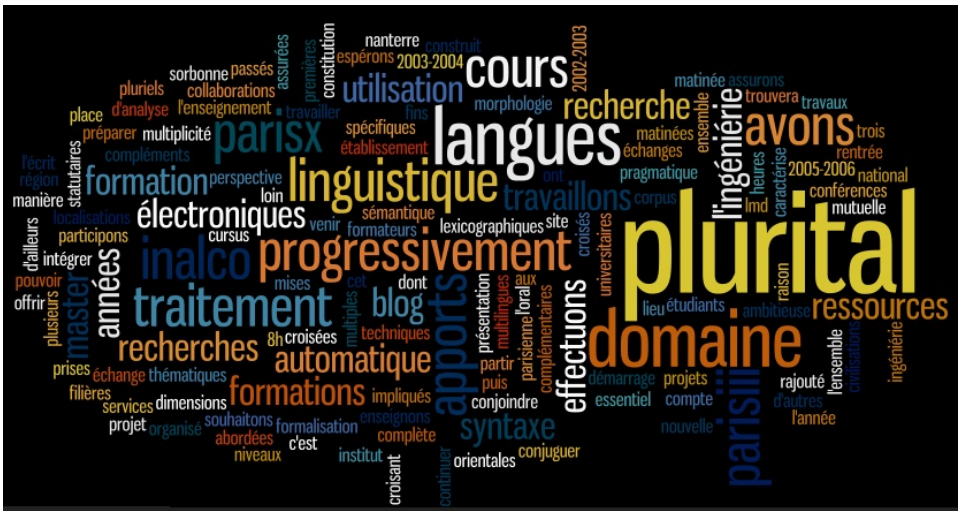
**Enseignant** : Sylvain Kahane (Paris Ouest Nanterre La Défense)
**Lieu** : Paris Ouest Nanterre La Défense
**Horaire** : vendredi 13h00-15h00

L'objectif est de présenter un modèle d'une langue naturelle, c'est-à-dire un dispositif permettant de simuler un sujet parlant, du sens qu'il souhaite communiquer au son qu'il produit (et notamment la prosodie). Nous aborderons la question des unités linguistiques élémentaires (morphèmes, unités lexicales, mots, constructions) et la question des différents types d'organisation de ces unités (organisation discursive et structure communicative, structure prédicat-argument, dépendance syntaxique, constituants topologiques, constituants prosodiques). Nous construirons ensemble un fragment de modèle pour le français et nous verrons comment lexique et grammaire s'articulent. Ce modèle s'inscrit dans le cadre des grammaires de dépendance et plus particulièrement de la Théorie Sens-Texte. Il emprunte aux grammaires lexicalisées le calcul de la structure d'un énoncé par la combinaison de structures élémentaires et aux grammaires d'unification le mode de combinaison de ces structures. Tous les outils mathématiques utilisés seront introduits et motivés par des questions théoriques.

**Bibliographie**

Bresnan Joan, 2001, *Lexical-Functional Syntax*, Blackwell.
Creissels Denis, 1995, *Eléments de syntaxe générale*, PUF.
Ducrot Oswald, 1995, Unités significatives, in Ducrot & Shaeffer, *Nouveau dictionnaire encyclopédique des sciences du langage*, Seuil.
Goldberg Adele, 1995, *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
Kahane Sylvain, 2002, *Grammaire d'Unification Sens-Texte : vers un modèle mathématique articulé de la langue*, Université Paris 7, 82 pages, www.kahane.fr.
Mel'čuk Igor, 1997, *Vers une linguistique Sens-Texte*, Leçon inaugurale au Collège de France, 78 p.
Polguère Alain, 2008, *Lexicologie et sémantique lexicale*, Presses de l'Université de Montréal
Sag Ivan, Thomas Wasow, Emily Bender, 2003, *Syntactic theory: A Formal Introduction*, CSLI Publications, Stanford.
Tesnière Lucien, 1959, *Eléments de syntaxe structurale*, Klincksieck.

---




# http://pluriTAL.org

## Contacts


### Cori Marcel

- Paris X
- MoDyCo
- mcori@u-paris10.fr

### Daube Jean-Michel

- INALCO
- CRIM
- jean-michel.daube@inalco.fr

### Fleury Serge

- Paris 3 / ILPGA
- SYLED/CLA$^2$T
- serge.fleury@univ-paris3.fr
- http://sfweb.no-ip.org.