

LICENCE LLCER

Parcours traitement numérique multilingue

2014 - 2015

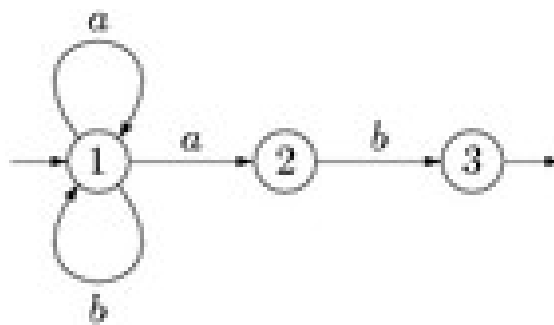


↑ 𐎎𐎗𐎚 национален שפה 文化 شرقية
i n a l c o

Institut national
des langues
et civilisations orientales

Sommaire

PRÉSENTATION GÉNÉRALE DU DÉPARTEMENT.....	1
CONTACT ET LIEUX D'ENSEIGNEMENT.....	1
ÉQUIPE PÉDAGOGIQUE.....	1
SECRÉTARIAT.....	1
ENSEIGNANTS.....	1
PUBLIC VISÉ.....	1
OBJECTIFS DU PARCOURS TNM.....	2
DÉBOUCHÉS DE LA FORMATION.....	2
ACCÈS AU PARCOURS TNM.....	2
LANGUES D'ACCÈS.....	2
LES PASSERELLES.....	3
POURSUITE D'ÉTUDES.....	3
STRUCTURE DE LA LICENCE LLCER, PARCOURS TNM.....	4
ORGANISATION DES ENSEIGNEMENTS.....	4
DESCRIPTION DES ENSEIGNEMENTS PROPRES AU PARCOURS TNM.....	5
ENSEIGNEMENTS DE L2.....	5
ENSEIGNEMENTS DE L3.....	7
INFORMATIONS PRATIQUES ET CALENDRIER.....	9
EMPLOIS DU TEMPS ET PLANNING DES EXAMENS.....	9
CALENDRIER UNIVERSITAIRE 2014-2015.....	9



Présentation générale du département

L'offre de formation du département TIM (LLCER, parcours Traitement Numérique Multilingue, Master Traitement Automatique des Langues) a pour objectif de former les étudiants en langue aux métiers de la société de l'information demandant de fortes compétences linguistiques.

À cette fin, elle leur permet de s'approprier les connaissances et les méthodes en traitement automatique des langues (TAL), domaine fondamentalement pluridisciplinaire (langue(s), linguistique, informatique théorique et pratique...)

Le TAL est de manière générale le domaine recouvrant toutes les applications informatiques ayant trait au « langage naturel » et suppose la modélisation de phénomènes langagiers afin de doter les systèmes informatiques de « compétences » dans une ou plusieurs langues pour analyser ou générer des énoncés dans ces langues. L'ingénierie de la langue est l'ensemble des produits et services qui ont le TAL pour objet.

Contact et lieux d'enseignement

Département Textes, Informatique, Multilinguisme



65, rue des Grands Moulins
75013 Paris

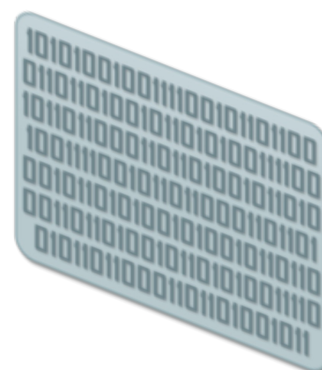


01 81 70 11 36



tnm@inalco.fr

Métro : ligne 14 Bibliothèque François Mitterrand



Équipe pédagogique

Secrétariat

Mme Amandine Polliart, bureau 324

Enseignants

Christine Bonnot	Professeur	linguistique
Egle Ensoo	Ingénieur	TAL
Christine Jacquet-Pfau	Maître de conférences	linguistique, TAL
M-Anne Moreaux	Maître de conférences	linguistique, informatique, TAL
Damien Nouvel	Maître de conférences	informatique, TAL
Patrick Paroubek	Ingénieur de recherche	informatique, TAL

Public visé

Le parcours Traitement Numérique Multilingue s'adresse à tout étudiant qui, à l'issue d'un L1 à l'INALCO, souhaite, dans l'optique de son insertion professionnelle, acquérir parallèlement à sa formation en langue les compétences théoriques et pratiques fondamentales du domaine du traitement automatique des langues et de l'ingénierie linguistique.

Objectifs du parcours TNM

Le parcours Traitement Numérique Multilingue de la licence est centrée sur l'acquisition des connaissances disciplinaires fondamentales du domaine et tient compte des prérequis d'entrée dans les différentes spécialités, professionnalisantes ou recherche, du Master Traitement Automatique des Langues.

Un spécialiste de TAL doit savoir constituer ses données linguistiques, les organiser et les modéliser, puis choisir ou élaborer une formalisation, enfin construire en conséquence une implémentation informatique et savoir l'évaluer. À ces fins, il doit posséder des connaissances et des compétences multiples : connaissance scientifique d'une langue (que cette langue soit une langue étrangère ou la langue maternelle) ; connaissances en sciences du langage et en linguistique descriptive ; connaissances fondamentales en mathématique et en logique ; connaissances en informatique théorique et pratique (systèmes, programmation en plusieurs langages) ; connaissances des notions théoriques communes à la linguistique et à l'informatique ; connaissances de plusieurs domaines d'application.

Débouchés de la Formation

- rédacteur de documentations techniques
- rédacteur de terminologie (multilingues)
- intervenant dans la constitution, l'organisation et la gestion de bases de données multilingues
- expert pour l'amélioration de la couverture linguistique d'applications existantes
- Participant à la conception et au développement d'outils d'ingénierie linguistique et de sites Web (TAO, EAO, recherche et filtrage d'informations, documentation automatique, outils bureautiques évolués, outils de veille technologique)

Les entreprises recrutant nos étudiants se trouvent dans les secteurs de l'édition (Larousse, Hachette, Robert,...), la traduction (Systran, Trados, Wipro...), de la veille (Publicis, Fircosoft, samestory...), de la gestion de projets de traduction (Lionbridge, Musée du Louvre), de l'édition (Le Robert, Larousse, Thomson Reuters...), de la structuration d'information (Rosebud, Arisem...), du e-learning (Accenture, Total...), de la gestion de sites web multilingues (Barclays...)

Accès au parcours TNM

Le parcours Traitement Numérique Multilingue est un parcours de L2-L3.

La première année de la licence (L1) est consacrée à l'étude de la langue dont l'étudiant veut devenir spécialiste et comporte différents enseignements généralistes. Le choix du parcours TNM se fait à l'entrée en deuxième année (L2).

Pour les étudiants de l'INALCO, il n'y a aucun prérequis pour s'inscrire à ce parcours au début de L2. Les étudiants venant d'autres établissements et souhaitant s'inscrire à ce parcours devront faire preuve de compétences équivalentes à celles du bloc « enseignements fondamentaux » de L1 dans la langue orientale dont ils sont spécialistes.

Langues d'accès

Le parcours TNM est accessible pour 64 des langues enseignées à l'INALCO :

albanais, amharique, arabe, arabe maghrébin-arabe oriental, arménien, bengali, berbère, birman, bosniaque-croate-serbe, bulgare, cambodgien, chinois, coréen, drehu, estonien, filipino, finnois, géorgien, grec moderne, haoussa, hébreu moderne, hindi, hongrois, indonésien-malais, japonais, laotien, letton, lituanien, macédonien, malgache, mandingue, mongol, népal, ourdou, pashto, persan, peul, polonais, roumain, rromani, russe, singhalais, slovaque, slovène, swahili, tahitien, tamoul, tchèque, thaï, tibétain, turc, ukrainien, vietnamien, wolof, et yoruba.

Les passerelles

Les étudiants souhaitant accéder au parcours TNM au niveau L3 sans avoir suivi le niveau L2 devront avoir validé tous les enseignements du bloc fondamental de L2. Ils devront suivre et valider tous les enseignements de L2 et L3, spécifiques au parcours TNM l'année de leur inscription.

Poursuite d'études

Les étudiants titulaires de la licence LLCER, parcours TNM sont admis de plein droit en master TAL (cohabilité avec Paris 3 et Paris 10)

M1 – tronc commun

M2 – 2 parcours professionnel (Ingénierie multilingue, Traductique et gestion de l'information)
et 1 parcours recherche (Recherche et développement)

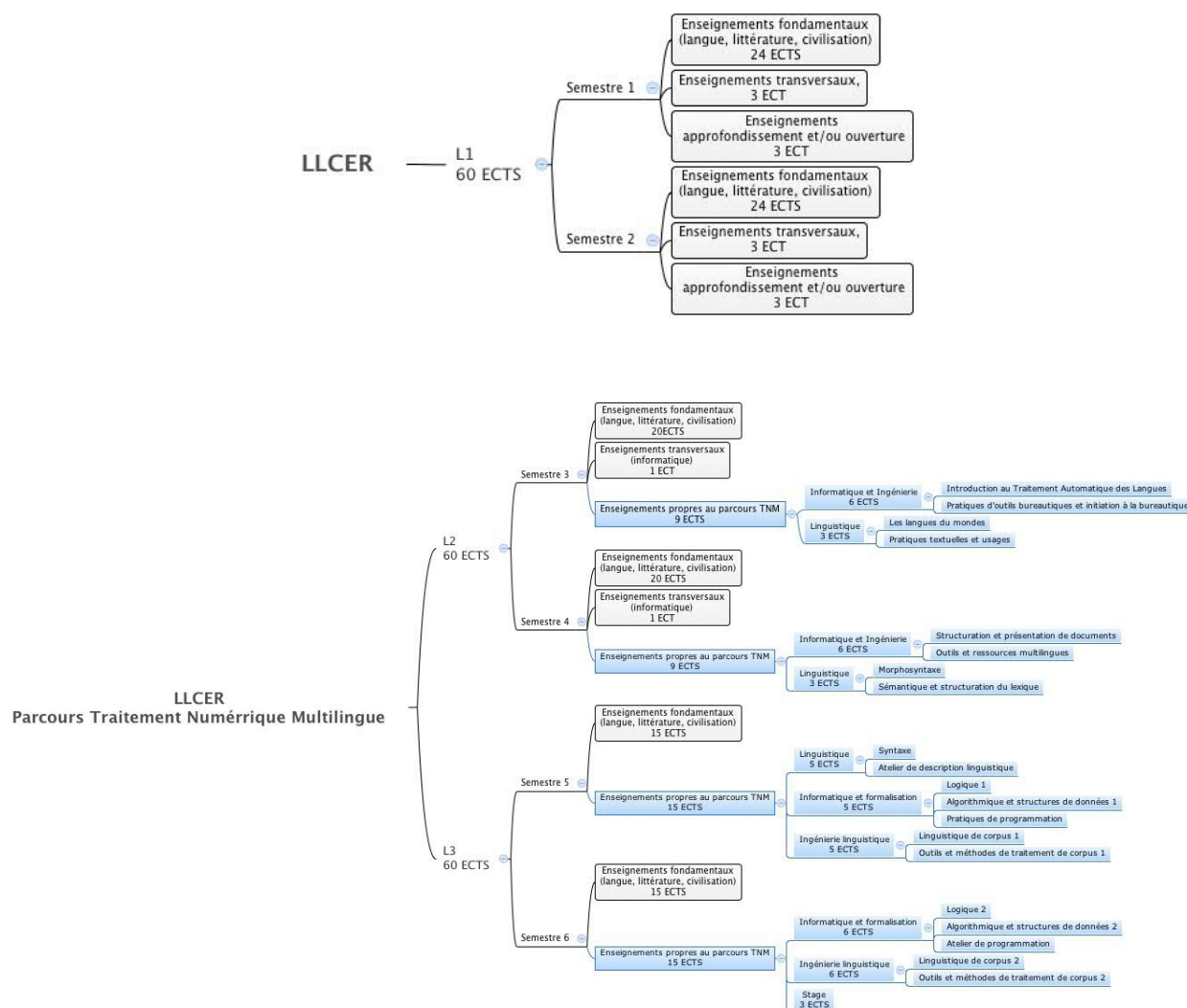


Structure de la licence LLCER, parcours TNM

Organisation des enseignements

La licence LLCER, parcours TNM est acquise en 180 ECTS : 30 ECTS par semestre. La première année, semestre 1 et 2, l'étudiant est inscrit en LLCE dans la langue qu'il souhaite étudier. À partir de la 2ème année, il continue l'étude de sa langue et suit les enseignements propres au parcours TNM.

Les enseignements se répartissent selon les schémas suivants :



Enseignements fondamentaux

Les enseignements fondamentaux (langue, littérature et civilisation) sont assurés par les départements de langue. La description de ces enseignements se trouve dans les brochures de langue.

Enseignements transversaux

- en L1, méthodologie et langue de travail
- en L2, informatique

Enseignements spécifiques au parcours TNM

Les enseignements spécifiques à la mention se répartissent sur L2 et L3 de façon homogène, à raison de 2 Unités d'Enseignement par semestre en L2 et de 3 unités d'enseignement par semestre en L3 :

Description des enseignements propres au parcours TNM

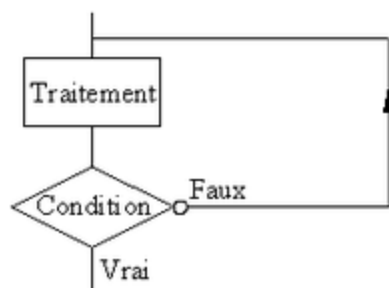
Enseignements de L2

L2-S1 (S3)	INFORMATIQUE ET INGÉNIERIE		5 ECTS
	<p>Introduction au TAL Cet enseignement a pour but de donner une introduction générale au traitement automatique de textes écrits et de faire prendre conscience aux étudiants des problèmes soulevés par le développement d'outils informatiques capables d'analyser ou de produire des textes rédigés dans une langue humaine. Les domaines d'application du TAL sont présentés puis un bref aperçu historique montre comment la discipline s'est constituée. On aborde ensuite les outils existants et les différentes connaissances linguistiques nécessaires pour traiter un texte qui n'est a priori pour la machine qu'une suite de caractères, donc de codes.</p>	CM	19h30
	<p>Pratiques d'outils bureautiques et initiation à la programmation L'objectif de ce cours est double. Il s'agit d'une part d'amener les étudiants à exploiter pleinement les possibilités des logiciels bureautiques courants (traitements de textes, tableurs, navigateurs) leur permettant de rechercher, structurer et formater leurs données linguistiques. D'autre part, à travers les environnements offerts par ces logiciels, le cours propose une initiation à la démarche algorithmiques, aux types de données et à la construction de programmes.</p>	TD	26h
	LINGUISTIQUE		4 ECTS
	<p>Les langues du monde (classification des langues, familles de langues, aires linguistiques) Le cours présentera d'abord les grandes familles de langues du point de vue de leurs apparentements génétiques, ainsi que de leur répartition géographique par continents (5 séances), puis examinera pour les nuancer à partir d'exemples les critères traditionnels de classification par grands types (isolant, fusionnel, agglutinant).</p>	CM	19h30
	<p>Pratiques textuelles et usages Les applications standard de l'analyse linguistique portent sur des textes (traduction, recherche d'information, dictionnaire). La notion de texte s'étend aujourd'hui à celle du document numérique et aux nouveaux usages qui y sont associés. L'objectif de ce cours est donc de faire acquérir une méthodologie d'analyse des données textuelles pour les opérations de caractérisation et d'extraction d'information.</p>	CM	19h30
			9 ECTS

XML
 Textes
 Morphologie
 Lexiques
 Documents
 HTML
 Semantique

L2-S2 (S4)	INFORMATIQUE ET INGÉNIERIE		5 ECTS
	Structuration et présentation de documents (langages de marquage : HTML, XML) Notions de caractère, texte, document, hyperdocument. Structuration implicite et explicite. Typographie, typographie numérique. Langages typographiques et langages de structuration, notions de balise. Présentation de html et xml. Utilisation de xml pour structurer un document sur des exemples concrets (introduction d'un certain nombre d'outils destinés à faciliter l'utilisation de xml.	CM : 13h TD : 13h	26h
	Outils et ressources multilingues L'objectif de ce cours est de présenter aux étudiants des outils (essentiellement correcteurs, et dictionnaires) qu'ils auront à utiliser à la fois comme pratiques mais également comme objets d'analyse. L'un des principaux aspects méthodologiques de ce cours consiste à développer une approche critique des outils existants : en comprendre le fonctionnement, en connaître les possibilités et les limites et à envisager des traitements linguistiques pour rendre plus satisfaisants certains de leurs outils.	CM : 9h30 TD : 10h	19h30
	LINGUISTIQUE		4 ECTS
	Morphosyntaxe Une phrase n'est pas un assemblage de mots isolés mais d'unités entretenant des relations et exerçant des fonctions les unes par rapport aux autres. On présentera et on analysera dans ce cours les notions fondamentales permettant de conduire les étudiants vers l'analyse syntaxique: unités de base de la description syntaxique (morphèmes, mots, syntagmes, phrases), catégories syntaxiques, fonctions syntaxiques, tests de segmentation, classification des types de phrases, structures superficielles et profonde. Ces notions seront illustrées sur différentes langues.	CM	19h30
	Sémantique et structuration du lexique Ce cours constitue une introduction à l'analyse des unités lexicales, telles que les consignent notamment les dictionnaires, lieu privilégié de structuration du lexique. Un dictionnaire véhicule en effet un savoir sur la construction du lexique, mais il est en même temps conçu selon un certain savoir linguistique : différents dictionnaires ne nous donnent pas le même type d'information sur le lexique et ne nous le présentent pas selon les mêmes niveaux d'analyse. Le cours abordera plus précisément des notions qui permettront aux étudiants de connaître les principales notions mises en œuvre dans la construction d'un lexique. Les thèmes seront abordés : du mot au morphème (définition, repérage, représentation dans les dictionnaires) ; du dictionnaire au texte, de la langue au discours ; repérage des informations sémantiques explicitement indiquées dans un dictionnaire (analyse du métalangage dans différents dictionnaires) ; analyse diachronique et analyse synchronique ; fonction de l'étymologie ; motivation et transparence.	CM	19h30
		9 ECTS	84h30

Total des 2 semestres de L2 pour la partie TNM	18 ECTS	169h
---	----------------	-------------



Enseignements de L3

L3-S1 (S5)	INFORMATIQUE ET FORMALISATION		5 ECTS
	<p>Logique 1 (logique des propositions et théorie des ensembles) Partant de la définition du traitement automatique des langues, on introduit la notion de langage formel et de sémantique compositionnelle. La logique des propositions est d'abord présentée comme un langage formel dont la sémantique (limitée) est compositionnelle. La représentation des formules sous forme d'arbres binaires et la détermination de leur valeur de vérité permettent d'illustrer ces notions sur un cas concret. La logique des propositions est ensuite appliquée au calcul : circuits logiques, machine à additionner, ... et à la notion de démonstration de la validité d'un argument, humaine (déduction naturelle) et automatique (algorithmique).</p>	CM : 13h TD : 13h	26h
	<p>Algorithmique et structures de données 1 Cet enseignement a pour but d'initier les étudiants aux principes de construction des algorithmes et de définition des structures de données. Après avoir présenté les notions de calcul et d'algorithme, le cours insiste sur les méthodes permettant de construire des solutions structurées (exploitation de l'énoncé pour dégager les tâches à accomplir et les objets à manipuler, analyse descendante, notions d'état et de configuration, notion de spécifications) et détaille les méthodes de gestion de certains types de données (données élémentaires, tableaux et chaînes de caractères).</p>	CM	13h
	<p>Pratiques de programmation Les notions algorithmiques présentées dans le cours Algorithmique et structures de données seront mises en œuvre à travers le langage C++.</p>	TD	26h
	LINGUISTIQUE		5 ECTS
	<p>Syntaxe En s'appuyant sur les notions présentées sur le cours « Morphosyntaxe générale », on présentera deux des principaux modèles de description des structures syntaxiques: d'une part les modèles dits en dépendance en s'appuyant sur la Syntaxe Structurale de Tesnière, d'autre part ceux en constituants immédiats à partir de la Grammaire Générative.</p>	CM	19h30
	<p>Atelier de description linguistique On soumettra aux étudiants des « problèmes de linguistique » thématiques, couvrant les différents niveaux de l'analyse linguistique, élaborés à partir de langues très diverses et généralement très peu connues, et destinés à les entraîner à examiner des données linguistiques inconnues de manière systématique. La maîtrise des notions linguistiques mises en jeu dans chaque exercice sera ainsi approfondie au fil du cours.</p>	atelier	19h30
	INGÉNIERIE LINGUISTIQUE		5 ECTS
	<p>Linguistique de corpus 1 Les quantités de textes aujourd'hui disponibles, en particulier sous forme électronique via internet sont vertigineuses. La forte demande sociale – en particulier les besoins des milieux professionnels – en matière d'indexation, d'extraction et d'analyse de contenu a suscité la mise au point de nouvelles applications et la création de nouveaux instruments de recherche entraînant de profondes mutations dans le travail des linguistes. L'objectif de ce cours, ainsi que celui de S6 est de montrer quelles sont les nouvelles perspectives et les nouvelles tâches qui s'offrent aux linguistes: comment constituer des ensembles cohérents de texte ? Comment les mettre sous une forme permettant leur exploitation ? Pour quelle application les soumet-on à l'analyse ? Quel type de traitement envisager et avec quels outils.</p>	CM	19h30
	<p>Outils et méthodes de traitement de corpus 1 Il s'agira de mettre en œuvre les notions présentées dans le cours <i>Linguistique de corpus</i>.</p>	TD	19h30
	15 ECTS	143h	

L3-S2 (S6)	INFORMATIQUE ET FORMALISATION		6 ECTS
	Logique 2 (logique des prédicats et démonstration) La logique des prédicats classiquement introduite par l'impossibilité de rendre compte de la validité de l'argument « si P → Q et si P alors Q » en logique des propositions. Les notions de théorie des ensembles étudiées dans le cadre du cours logique 1 sont utilisées pour introduire les notions fondamentales de la logique des prédicats du premier ordre : domaine, sémantique des prédicats, quantification, sémantique des expressions. La notion de relation est ensuite abordée, puis celle de relation d'équivalence et de partition d'un ensemble, et enfin celle de relation d'ordre total ou partiel, illustrée notamment par l'étude formelle de l'ordre lexicographique.	CM : 13h TD : 13h	26h
	Algorithmique et structures de données 2 Ce cours insiste sur la structuration des données et aborde la gestion de structures complexes et l'algorithmique correspondante (listes, arbres) pour aboutir aux méthodes de construction de types propres au domaine du traitement des langues.	CM	13h
	Atelier de programmation Il s'agira de suivre et de guider la progression du travail des étudiants ayant à réaliser un projet dont le sujet sera déterminé par l'ensemble des enseignants. L'objet de ce travail sera d'analyser un problème de langue, de le décrire, d'en présenter une solution formalisée et programmée. Il donnera lieu à un petit mémoire présenté devant jury. Il a pour objectif de montrer la capacité des étudiants à poser un problème linguistique et à y répondre en utilisant des méthodes et des outils issus de l'informatique.	atelier	26
	INGÉNIERIE LINGUISTIQUE		6 ECTS
	Linguistique de corpus 2 voir Linguistique de corpus 1.	CM	19h30
	Outils et méthodes de traitement de corpus 2 voir Outils et méthodes de traitement de corpus 1.	TD	19h30
	STAGE		3 ECTS
		stage	1 mois
		15 ECTS	104h
Total des 2 semestres de L3 pour la partie TNM		30 ECTS	247h
Total L2 et L3 pour la partie TNM		48 ECTS	416h

Syntaxe
 Automates
 Indexation
 Logique
 C++
 Algorithmique
 Corpus
 Programmation

Informations pratiques et calendrier

Emplois du temps et planning des examens

Les cours propres au TNM ont lieu pour :

- L2 : le mercredi après-midi et le vendredi matin
- L3 : le lundi après-midi et le mercredi matin

Pour chaque enseignement, sauf pour ceux de linguistique, le dernier partiel ou l'examen final a lieu durant la dernière séance de cours. Pour les enseignements de linguistique, les contrôles se déroulent dans la plage réservée aux enseignements transversaux.

Le planning des cours et des examens est affiché sur le site de l'INALCO.

Calendrier universitaire 2014-2015

Premier semestre : du lundi 22 septembre 2014 au samedi 10 janvier 2015

Congés automne : du lundi 27 octobre au samedi 1er novembre 2014

Congés hiver : du lundi 22 décembre 2014 au samedi 3 janvier 2015

Examens 1^{ère} session du 1^{er} semestre : du lundi 12 janvier au samedi 24 janvier 2015

Second semestre : du lundi 26 janvier au samedi 16 mai 2015

Congés hiver : du lundi 23 février au samedi 28 février 2015

Congés printemps : du lundi 20 avril au samedi 2 mai 2015

Examens 1^{ère} session du 2nd semestre : lundi 18 mai au samedi 30 mai 2015

Examens 2^{ème} session : lundi 15 juin au samedi 4 juillet 2015

Stage

Un stage d'un mois minimum est obligatoire en L3. Des précisions concernant ce stage vous seront données lors de la première semaine de cours.

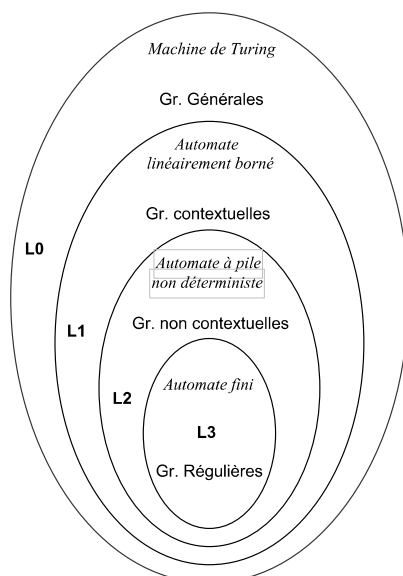


Illustration 1: « Hiérarchie Chomsky » par Jean-Christophe BENOIST — Travail personnel. Sous licence Creative Commons Attribution-Share Alike 3.0 via Wikimedia Commons - <http://commons.wikimedia.org/wiki/File:Hi%C3%A9rarchieChomsky.svg#mediaviewer/Fichier:Hi%C3%A9rarchieChomsky.svg>

A word cloud of French terms related to linguistics and computational linguistics. The words are arranged in a roughly rectangular shape, with varying sizes and orientations. The largest words are 'Grammaires' and 'Licence', both oriented vertically. Other prominent words include 'Dictionnaires', 'Langues', 'Informatique', 'Textes', 'Indexation', 'Fouille', 'Multilingue', 'Web', 'Numérique', 'Master', 'Linguistique', 'TAL', 'TAO', 'INALCO', 'Traitement', and 'TIM'. The colors range from light blue to a darker blue.

TIM
Multilingue
Fouille
Indexation
INALCO
TAO
Textes
Dictionnaires
Langues
Master
Linguistique
TAL
Informatique
Web
Numérique
Traitement
Licence
Grammaires