

5.0 crédits	30.0 h + 15.0 h	2q
-------------	-----------------	----

Enseignants:	Dupont Pierre ; Fairon Cédric ;
Langue d'enseignement:	Anglais
Lieu du cours	Louvain-la-Neuve
Préalables :	-- Algorithmique et de préférence des connaissances de base en apprentissage automatiques (tels que visés par les cours SINF1121 et INGI2262)
Thèmes abordés :	-- Concepts de base en phonologie, morphologie, syntaxe et sémantique -- Ressources linguistiques -- Etiquetage en partie du discours -- Modélisation statistique de la langue (N-grams et modèles de Markov cachés) -- Algorithmes d'analyse robuste, grammaires hors-contexte probabilistes -- Applications de l'ingénierie linguistique telles que les logiciels de correction orthographique ou grammaticale, l'étiquetage en partie du discours ,l'indexation automatique de documents ou la catégorisation de textes
Acquis d'apprentissage	Les étudiants ayant suivi avec fruit ce cours seront capables de -- décrire les concepts fondamentaux de la modélisation du langage naturel -- maîtriser la méthodologie de l'utilisation de ressources linguistiques (corpus, dictionnaires, réseaux sémantiques, etc) et faire un choix argumenté entre les différentes ressources linguistiques -- appliquer de manière pertinente les techniques statistiques de modélisation du langage -- développer des applications en ingénierie linguistique Les étudiants auront développé des compétences méthodologiques et opérationnelles. En particulier, ils auront développé leur capacité à -- s'intégrer dans une approche pluridisciplinaire à la frontière entre l'informatique et la linguistique, en utilisant à bon escient la terminologie et les outils de l'une ou l'autre discipline, -- gérer au mieux le temps disponible pour mener à bien des mini-projets, -- manipuler et exploiter de grandes quantités de données. <i>La contribution de cette UE au développement et à la maîtrise des compétences et acquis du (des) programme(s) est accessible à la fin de cette fiche, dans la partie « Programmes/formations proposant cette unité d'enseignement (UE) ».</i>
Modes d'évaluation des acquis des étudiants :	25 % pour les travaux pratiques + 75 % examen final (livre fermé) Pas de possibilité de refaire les travaux en seconde session
Méthodes d'enseignement :	-- 12 cours magistraux -- 3 miniprojets -- séance de discussion sur les corrections des miniprojets
Contenu :	-- Bases en linguistique : morphologie, partie du discours, structures de phrase, sémantique et pragmatique -- Rappels mathématiques : langages formels et éléments de la théorie de l'information -- Analyse de corpus : formatage, tokenization, morphologie, marquage des données -- N-grams : estimation par maximum de vraisemblance et lissage --

	<p>Modèles de Markov Cachés : définitions, algorithmes de Baum-Welch et de Viterbi -- Etiquetage en parties du discours -- Grammaires hors-contexte probabilistes : estimation des paramètres et algorithmes d'analyse, utilisation de treebanks -- Traduction automatique: méthodes classiques et statistiques (modèles IBM, modèles basés sur la phrase), évaluation -- Applications : prédicteur de mots pour SMS, marquage de parties de texte, outils d'extraction d'information http://www.icampus.ucl.ac.be/claroline/course/index.php?cid=INGI2263</p>
<p>Bibliographie :</p>	<p>Slides obligatoires disponibles sur le site : http://www.icampus.ucl.ac.be/claroline/course/index.php?cid=INGI2263 1 textbook conseillé : -- Speech and Language Processing (2nd Edition), D. Jurafsky and J.H. Martin, Prentice Hall, 2009.</p>
<p>Cycle et année d'étude :</p>	<p>> Master [120] en linguistique > Master [120] : ingénieur civil en informatique > Master [120] en sciences informatiques > Master [120] en statistiques, orientation générale</p>
<p>Faculté ou entité en charge:</p>	<p>INFO</p>