






5 crédits	30.0 h + 15.0 h	Q1
-----------	-----------------	----

Enseignants	Dupont Pierre ;Fairol Cédrick ;
Langue d'enseignement	Anglais
Lieu du cours	Louvain-la-Neuve
Thèmes abordés	<ul style="list-style-type: none"> <li>• Concepts de base en phonologie, morphologie, syntaxe et sémantique</li> <li>• Ressources linguistiques</li> <li>• Etiquetage en partie du discours</li> <li>• Modélisation statistique de la langue (N-grams et modèles de Markov cachés)</li> <li>• Algorithmes d'analyse robuste, grammaires hors-contexte probabilistes</li> <li>• 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</li> </ul>
Acquis d'apprentissage	<p>Eu égard au référentiel AA du programme « Master ingénieur civil en informatique », ce cours contribue au développement, à l'acquisition et à l'évaluation des acquis d'apprentissage suivants :</p> <ul style="list-style-type: none"> <li>• INFO1.1-3</li> <li>• INFO2.3-4</li> <li>• INFO5.3-5</li> <li>• INFO6.1, INFO6.4</li> </ul> <p>Eu égard au référentiel AA du programme « Master [120] en sciences informatiques », ce cours contribue au développement, à l'acquisition et à l'évaluation des acquis d'apprentissage suivants :</p> <ul style="list-style-type: none"> <li>• SINF1.M4</li> <li>• SINF2.3-4</li> <li>• SINF5.3-5</li> <li>• SINF6.1, SINF6.4</li> </ul> <p>1</p> <p>Les étudiants ayant suivi avec fruit ce cours seront capables de</p> <ul style="list-style-type: none"> <li>• décrire les concepts fondamentaux de la modélisation du langage naturel</li> <li>• maîtriser la méthodologie de l'utilisation de ressources linguistiques (corpus, dictionnaires, réseaux sémantiques, etc) et faire un choix argumenté entre les différentes ressources linguistiques</li> <li>• appliquer de manière pertinente les techniques statistiques de modélisation du langage</li> <li>• développer des applications en ingénierie linguistique</li> </ul> <p>Les étudiants auront développé des compétences méthodologiques et opérationnelles. En particulier, ils auront développé leur capacité à</p> <ul style="list-style-type: none"> <li>• 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,</li> <li>• gérer au mieux le temps disponible pour mener à bien des mini-projets,</li> <li>• manipuler et exploiter de grandes quantités de données.</li> </ul> <p>-----</p> <p><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></p>
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	<ul style="list-style-type: none"> <li>• 12 cours magistraux</li> <li>• 3 miniprojets</li> <li>• séance de discussion sur les corrections des miniprojets</li> </ul>
Contenu	<ul style="list-style-type: none"> <li>• Bases en linguistique : morphologie, partie du discours, structures de phrase, sémantique et pragmatique</li> <li>• Rappels mathématiques : langages formels et éléments de la théorie de l'information</li> <li>• Analyse de corpus : formatage, tokenization, morphologie, marquage des données</li> <li>• N-grams : estimation par maximum de vraisemblance et lissage</li> <li>• Modèles de Markov Cachés : définitions, algorithmes de Baum-Welch et de Viterbi</li> <li>• Etiquetage en parties du discours</li> </ul>

	<ul style="list-style-type: none"><li>• Grammaires hors-contexte probabilistes : estimation des paramètres et algorithmes d'analyse, utilisation de treebanks</li><li>• Traduction automatique: méthodes classiques et statistiques (modèles IBM, modèles basés sur des segments de phrases), évaluation</li><li>• Applications : prédicteur de mots pour SMS, marquage de parties de texte, outils d'extraction d'information <a href="http://moodleucl.uclouvain.be/course/view.php?id=7865">http://moodleucl.uclouvain.be/course/view.php?id=7865</a></li></ul>
Ressources en ligne	<a href="http://moodleucl.uclouvain.be/course/view.php?id=7865">http://moodleucl.uclouvain.be/course/view.php?id=7865</a>
Bibliographie	Slides obligatoires disponibles sur le site : <a href="http://moodleucl.uclouvain.be/course/view.php?id=7865">http://moodleucl.uclouvain.be/course/view.php?id=7865</a> 1 textbook conseillé : <ul style="list-style-type: none"><li>• <a href="#">Speech and Language Processing (2nd Edition)</a>, D. Jurafsky and J.H. Martin, Prentice Hall, 2009.</li></ul>
Faculté ou entité en charge:	INFO

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] : ingénieur civil en science des données	DATE2M	5		
Master [120] : ingénieur civil en informatique	INFO2M	5		
Master [120] en linguistique	LING2M	5		
Master [120] en sciences informatiques	SINF2M	5		
Master [120] en science des données, orientation statistique	DATS2M	5		
Master [120] en science des données, orientation technologies de l'information	DATI2M	5		