

LSV 727 LANGUE ET NOUVELLES TECHNOLOGIES
(3 crédits)

PRÉALABLE : aucun

Nombre limite : 25 étudiantes ou étudiants

DESCRIPTEURS :

Objectifs : Évaluer les avantages, les problèmes et les limites liés au traitement informatique des langues naturelles, qu'il s'agisse d'analyse ou de génération; poser un regard critique sur les ressources et les outils orientés vers le traitement des données langagières.

Contenu : Réflexion sur l'histoire et les principes généraux du traitement automatique des langues. Examen exhaustif des ressources et des outils informatiques en lien avec le traitement du français (bases de données lexicales et terminologiques, banques de données textuelles, dictionnaires unilingues et bilingues, logiciels d'aide à la transcription de l'oral, à la révision linguistique et à la traduction).

AUTRES INFORMATIONS GÉNÉRALES

Pour vérifier si cette activité pédagogique est offerte à la présente session, consulter l'horaire sur le site du Département des arts, langues et littératures à la rubrique Étudiants actuels (www.usherbrooke.ca/dall/etudiants-actuels).

Précisions sur le contenu

Ce cours abordera les grandes thématiques et questions suivantes :

- Histoire des domaines du traitement automatique du langage (TAL) et de l'ingénierie linguistique
- Corpus et bases de données disponibles pour la linguistique
- Outils disponibles pour le traitement et l'analyse de données linguistiques
- Types de formats pour les données linguistiques (texte brut, HTML, XML, JSON)
- Extraction et manipulation de motifs dans des données linguistiques
- Traduction automatique
- Le traitement automatique de la parole (analyse du signal, reconnaissance et synthèse de la parole)
- Les systèmes d'intelligence artificielle générative
- L'influence de considérations pratiques (ex : ressources disponibles, coût de développement) sur l'architecture de systèmes de TAL
- Les apports et les limites de l'apprentissage automatique appliqué au langage
- Le rôle et la place des théories linguistiques dans la conception de systèmes de TAL.

Approche pédagogique

- cours magistraux entremêlés de discussions : il est attendu des étudiantes et étudiants une participation active à ces discussions tout au long du trimestre.
- ateliers pratiques : nous consacrerons une partie de chaque séance à la manipulation d'outils et/ou de données linguistiques.
- lectures complémentaires : les cours seront accompagnés de lectures d'article ou de chapitres d'ouvrage qui viendront compléter et/ou approfondir les points vus dans le cours.

Évaluation

L'évaluation de ce travail repose sur trois composantes :

- Exercices pratiques (en continu tout au long du trimestre) : 50%
- Examen en classe : 30%
- Présentation orale (en binôme) : 20%

Bibliographie sélective

Une bibliographie détaillée sera fournie au début du trimestre.

André, Virginie, Christophe Benzitoun, Emmanuelle Canut, Jeanne-Marie Debaisieux, Bertrand Gaiffe & Evelyne Jacquy. 2010. Traitement informatique de données orales: quels outils pour quelles analyses? *Les actes* 131.

Anthony, Laurence. 2004. AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. *proceedings of IWLeL*. Tokyo 7–13.

Boersma, Paul. 2014. The Use of Praat in Corpus Research. In Jacques Durand, Ulrike Gut & Gjert Kristoffersen (eds.). *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press, 342–360.

- Dalbera, Jean-Philippe. 2002. Le corpus entre données, analyse et théorie. *Corpus. Bases, corpus et langage* - UMR 6039 (1). <https://doi.org/10.4000/corpus.10>.
- Dutoit, Thierry. 2001. *An Introduction to Text-to-Speech Synthesis*. Softcover reprint of the original 1st ed. 1997. Dordrecht: Kluwer Academic Publishers.
- Dutoit, Thierry, Laurent Couvreur, Fabrice Malfrère, Vincent Pagel & Christophe Ris. 2002. Synthèse vocale et reconnaissance de la parole: Droites gauches et mondes parallèles. *Actes du 6è Congrès Français d'Acoustique*.
- Haton, Jean-Paul, Christophe Cerisara, Dominique Fohr, Yves Laprie & Kamel Smaïli. 2006. *Reconnaissance automatique de la parole - Du signal à son interprétation*. Paris: Dunod.
- Heba, Abdelwahab. 2021. *Reconnaissance automatique de la parole à large vocabulaire: des approches hybrides aux approches End-to-End*. Université Paul Sabatier - Toulouse III Thèse de doctorat non publiée. <https://tel.archives-ouvertes.fr/tel-03616588>. (22 August, 2022).
- Kurdi, Mohamed Zakaria. 2017. *Traitement automatique des langues et linguistique informatique 1*. Sciences cognitives édition. London: ISTE Editions.
- Kurdi, Mohamed Zakaria. 2018. *Traitement automatique des langues et linguistique informatique 2*. London: ISTE Editions.
- Pierrel, Jean-Marie. & Jean-Marie Pierrel. 2000. *Ingénierie des langues* (Traité IC2 Information, commande, communication. Série Informatique et systèmes d'information). Paris: Hermès Science publications. <http://catalogue.bnf.fr/ark:/12148/cb371950206>.
- Pincemin, Bénédicte. 2006. Concordances et concordanciers: de l'art du bon KWAC. In Rastier, François; Ballabriga, Michel; Duteil-Mougel, Carine; Fouquié, & Baptiste; (eds.), *XVIIe colloque d'Albi Lagages et signification - Corpus en Lettres et Sciences sociales: des documents numériques à l'interprétation*, 33–42. Albi, France: CALS-CPST. <https://halshs.archives-ouvertes.fr/halshs-00356008>. (22 August, 2022).
- Pincemin, Bénédicte. 2020. La textométrie en question. *Le Français Moderne - Revue de linguistique Française* 88(1). 26-43.
- Poibeau, Thierry. 2017. *Machine translation*. Cambridge (É.-U.) et Londres (R.-U.): MIT Press.