

Emmanuel Cartier

Maître de conférences en linguistique informatique à l'université Paris 13 Sorbonne Paris Cité depuis septembre 2006
membre du LIPN UMR 7030, équipe Représentation des Connaissances et Langage Naturel (RCLN) depuis 2015.
Membre du Labex *Empirical Foundations of Linguistics* (EFL)
Co-responsable axe « Formations » du Labex EFL depuis 2015 puis Responsable depuis 2018.
Coordinateur du projet Néoveille depuis 2015.
emmanuel.cartier@lipn.univ-paris13.fr

Parcours professionnel

2006-	Maître de Conférences, Université PARIS 13 SPC, Section 7
2004–2006	e-doc Labs, Paris, France, Directeur scientifique
1999–2004	LexiQuest puis Lingway, Montreuil, France, Ingénieur linguiste informaticien puis chef de projet et associé
1998–1999	Softissimo, Paris, France, Ingénieur linguiste informaticien
1996–1998	CNRS, Paris, France, Ingénieur Contractuel (Projet SERAPHIN, Résumé automatique)

Formation

2018	Habilitation à Diriger des Recherches , Université Paris 13 SPC, Dynamisme lexical des langues : éléments théoriques, méthodes automatiques, expérimentations en français contemporain soutenue le 13/12/2018, université Paris 13 Sorbonne Paris Cité. <u>Jury :</u> Jean-François Sablayrolles, Professeur émérite, Université Paris 3 Sorbonne Paris Cité, garant, section 7 Richard Huyghe, Professeur, Université de Fribourg (Suisse), rapporteur, section 7 Dominique Legallois, Professeur, Université Paris 3 Sorbonne Nouvelle, rapporteur, section 7 Alessandro Lenci, Professeur, Université de Pise (Italie), rapporteur, section 7 Esme Winter-Froemel, Professeur, Université de Trier (Allemagne), rapporteur, sections 7 et 17 Viviane Arigne, Professeur, Université Paris 13 Sorbonne Paris Cité, présidente, sections 7 et 11 Thierry Charnois, Professeur, Université Paris 13 Sorbonne Paris Cité, section 27.
2004	Doctorat en Linguistique informatique , Université Paris-IV Sorbonne, mention Très honorable Repérage automatique des expressions définitoires : modélisation de l'information définitoire. méthode d'exploration contextuelle, méthodologie du développement des ressources linguistiques, description des expressions du français contemporain, implémentation informatique Sous la direction de Jean-Pierre Desclés professeur, Paris IV Sorbonne

1993 **D.E.A. Mathématiques, Informatique, Linguistique** □
Université Paris-IV Sorbonne, mention Très Bien, major de promotion □
Description formelle des verbes de mouvement du français contemporain à partir du modèle de Boguslavski et Bogacki

Activités de recherche

Thématiques de recherche

Linguistique française et générale : lexicologie, morphologie, sémantique, constructions, description syntaxico-sémantique, néologie et changements sémantiques

Traitement automatique des langues / Linguistique de corpus / Humanités numériques : analyse automatique des textes, linguistique de corpus, approches distributionnelles et statistiques, modélisation des phénomènes linguistiques, expressions polylexicales, constructions, relations sémantiques.

Projets de recherche

2019-2020 Coordinateur projets PHC Barrande et Polonium : Emprunts néologiques en français, tchèque et polonais. Etude outillée sur corpus. En collaboration avec l'université Charles de Prague et l'université de Lodz.

2015-2018 Coordinateur du projet « NEOVEILLE » projet de recherche IDEX SPC, Repérage, analyse et suivi des néologismes en onze langues sur corpus dynamique (LIPN, HTL, CLILLAC-ARP, Ertim, Université Sao Paolo, groupe EMPNEO). www.neoveille.org

2017-2018 Coordinateur du projet Néonaute, moteur de recherche pour l'exploration d'informations linguistiques dans les archives web de la Bibliothèque Nationale de France (BnF), partenariat LIPN-BnF-Université de Starsbourg. <http://tal.lipn.univ-paris13.fr/neonaute/html/index.php>

2010-2013 Responsable scientifique du projet « Séquences verbales figées » contrat de recherche LDI-ITESOFT. Co-direction de la thèse CIFRE d'Aurélié Joseph (thèse soutenue en décembre 2013, mention très bien avec félicitations du jury).

2007-2014 Projet NEOLOGIA : base de données pour la gestion des bases de données lexicographiques sur les néologismes en français. Responsable de la modélisation et de l'implémentation informatique de la base de données. En collaboration avec Jean-François Sablayrolles. LDI

2002-2003 Projet Européen « **CROSSMARC** » d'extraction automatique d'informations dans les Offres d'emploi et les annonces de ventes de produits informatiques. Collaboration internationale entre Institute of Informatics & Telecommunications, NCSR « Demokritos » (GR), University of Edinburgh (UK), University Roma, Tor vergata (IT), Lingway (FR). Responsable de la partie française et de la méthodologie d'extraction des informations.

2000-2001 Projet Européen « **MUSI** » (Multi-lingual Summarisation Tool for the Internet) de résumé multilingue. Collaboration Internationale entre DFKI, Language Technology Lab (GER), University of Pisa, Instituto di Linguistica Computazionale (IT), Lexiquest (FR). Responsable de la partie française du

projet et de la méthodologie d'extraction des « phrases importantes ».

Encadrement de thèses de doctorat et de mémoires de master

Co-encadrement thèses de doctorat

En tant que MCF, j'ai été co-encadrant à 3 reprises, dont 2 en cours.

Thèses en cours :

Valeriya Vinogradova, *La créativité lexicale en français contemporain : troncation, autres mécanismes néologiques, traitement informatique des données*, inscription en septembre 2015, Directeur : Gabrielle Le Tallec, contrat doctoral de l'école doctorale ERASME.

Анна Мороз, *Neologismos contemporáneos en ruso y español : investigaciones estadísticas y lingüísticas*, inscription septembre 2018, Directeur : Pedro Mogorron Huerta, Universidad de Alicante.

Thèse soutenues :

Aurélie Joseph, décembre 2013, Repérage automatique de séquences figées, sous la direction de Salah Mejri et Emmanuel Cartier, thèse CIFRE (Entreprise : ITESOFT).
Mention très bien avec les félicitations du jury à l'unanimité.

Encadrement de mémoires de master

Dans le cadre de ma participation aux Masters SLATEL (M1 puis M1 et M2 depuis 2016) et TILDE (M2, depuis 2006), j'ai encadré en moyenne entre 7 à 5 étudiants en Master 1, et 3 étudiants en M2. Ci-dessous un échantillon des mémoires qui me semblent les plus significatifs, le titre et la mention obtenue.

2011-2012	5 M1, 5 M2 TILDE	Paola Carrion Gonzalez , Traitement automatique des langues créoles à base lexicale française (M2 TILDE, mention Très bien), Wilson Fred , Outils d'apprentissage automatique de ressources linguistiques à partir de corpus (M2 TILDE, Mention Très bien), Lamis Mhedhbi , Etude des séquences figées pour la catégorisation de documents (M2 TILDE, mention très bien), Caroline Langlet , Modalités et structures discursives pour l'Extraction d'Information (M2 TILDE, mention très bien), Meritxell Fernandes , Création et gestion semi-automatique de dictionnaire terminologique juridique (FR, EN, ES) à partir de corpus (M2 TILDE, Mention Très bien).
2013-2014	6 M1, 3 M2 TILDE	Ola El Ghamry , Détection d'entités nommées en arabe égyptien (M2 TILDE, mention Bien).
2014-2015	5 M1, 2 M2 TILDE	Elise Moreau , la langue des signes française : histoire, modèles linguistiques et état de l'art en TAL (M1 TAL, mention Très bien).
2015-2016	5 M1,	Laetitia Haroutunian , Classification automatique des opinions

3 M2 TILDE	<p>sur les réseaux sociaux : intérêt pour l'évaluation de campagnes de santé publique et application pratique autour de la websérie « Puceaux » (M2 TILDE, mention Bien),</p> <p>Dalia Saigh, Les sens verbaux, du lexique au corpus. Analyse de quelques verbes du français (M2 TILDE, mention Bien),</p> <p>Anne-Laure Piat-Marchand, Reconnaissance des entités nommées, exploration de différentes méthodes, application à un corpus de presse française(M2 TILDE, mention Bien).</p>
2016-2017 5 M1, 1 M2 COLEDI 2 M2 TILDE	<p>Amazigh Zafani, néologismes contemporains dans la presse quotidienne francophone algérienne (M1 TAL, mention Bien),</p> <p>Ibrahimera Timera, néologismes contemporains dans la presse quotidienne francophone du Sénégal (M1 TAL, mention Bien)</p> <p>Massimo Bertocci, Etude de vingt verbes polysémiques du français contemporain en vue d'une modélisation informatique, (M2 COLEDI, mention très bien).</p>
2017-2018 7 M1, 2 M2 TILDE	<p>Amazigh Zafani, Reconnaissance automatique de relations économiques entre entreprises dans des corpus économiques (M2 TILDE, soutenance fin octobre 2018)</p> <p>Ibrahimera Timera, Identification des courriels avec Nooj, expérimentation en entreprise, (M2 TILDE, soutenance fin octobre 2018)</p>

Autres expériences d'encadrement

Par ailleurs, dans le cadre du Master 1 en Informatique à Paris 13, j'ai été amené à encadrer à deux reprises des stagiaires durant 4 mois pour travailler respectivement sur la récupération automatique de corpus à partir de fils RSS sur Internet (projet Néoveille), et sur la mise en œuvre d'outils de génération de modèles de langue (Word Embeddings) pour la détection de néologie sémantique (projets Néoveille et Néonaute).

Dans le cadre du projet Néoveille (2015-2018), j'ai encadré un ingénieur d'études (Gaël Lejeune) pendant deux ans du projet, ainsi qu'une post-doctorante (Kata Gabor).

Dans le cadre du projet Néonaute (2017-2018), j'ai encadré un ingénieur d'études (Loïc Galand) pendant la durée du projet, et co-encadré un stagiaire niveau M2(Nicolas Scarcella).

Activités d'enseignement

Enseignements de linguistique générale et française

2010-2018	Introduction aux sciences du langage	Licence Lettres
-----------	--------------------------------------	-----------------

		modernes (L1), U. Paris 13, 18h / an
2014-2018	Introduction à la grammaire française : la phrase simple	Licence Lettres modernes (L1), U. Paris 13, 18h / an
2014-2018	Introduction à la grammaire française : la phrase complexe	Licence Lettres modernes (L1), U. Paris 13, 18h / an
2010-2014	Savoir dire, savoir écrire	Licence Lettres modernes (L2), U. Paris 13, 18h / an
2008-2012	Ecrits professionnels	Licence Lettres modernes (L2), U. Paris 13, 18h / an
2016-2018	Problématiques de la néologie	Master 2 COLEDI, U. Paris 13, 24h/an

Enseignements de linguistique informatique

2008-2016	Bases de données et lexicographie	Licence Lettres modernes (L2), U. Paris 13, 18h / an
2008-2016	Document numérique : analyse structurelle	Licence Lettres modernes (L2-L3), U. Paris 13, 18h / an
2008-2018	initiation au TAL	Licence Lettres modernes (L3), U. Paris 13, 18h / an
2014-2018	Outils informatiques pour la traduction	Licence Lettres modernes / LLCE (L3), U. Paris 13, 18h / an
2006-2018	Introduction au TAL	Master 1 TAL/SDL, U. Paris 13, 24h/an
2006-2014	Programmation en PERL (puis PYTHON) pour linguistes	Master 2 TILDE, U. Paris 13, 30h/an
2006-2018	Analyse automatique des textes en Python	Master 2 TILDE, U. Paris 13, 32h/an

Enseignements de linguistique

Les enseignements de linguistique dispensés portent à la fois sur les concepts fondamentaux de la linguistique générale, et sur les concepts pour l'analyse du français contemporain, centrés sur les notions de la grammaire : phrase simple et phrase complexe. Depuis 2016, j'assure également un séminaire de master 2 COLEDI sur la néologie. A cela s'ajoutent deux

enseignements en licence Lettres modernes (L2), non détaillés ici, sur : *savoir dire, savoir écrire* (2010-2014, sur la présentation orale du parcours professionnel et d'un sujet spécifique ; sur la rédaction d'un résumé de document), et *écrits professionnels* (2008-2012, rédaction d'un CV, d'une lettre de motivation).

Introduction aux Sciences du langage (18h CM/TD, Licence Lettres Modernes L1S1)

Objectifs du cours	Faire acquérir aux étudiants une culture minimale en Sciences du langage, sur les trois années de la licence. Ce cours est le premier cours, pour les licences 1.
Contenu du cours	<p>Présentation générale de la discipline :</p> <ul style="list-style-type: none"> - rapide historique de la linguistique - linguistique interne / externe - champs de la linguistique - domaines connexes : sociologie, psychologie, neurologie - linguistique théorique / linguistique appliquée <p>Linguistique historique / grammaires comparatives :</p> <ul style="list-style-type: none"> - Histoire de l'écriture - Génétique des langues - Typologie des langues <p>Linguistique structurale : Ferdinand de Saussure -</p> <ul style="list-style-type: none"> - concepts fondateurs 1 : langage / langue(s) / parole, diachronie / synchronie, linguistique / sémiologie, la langue comme système, objet(s) de la linguistique - concepts fondateurs 2 : notion de signe linguistique (signifiant / signifié, propriétés du signe, notion de valeur, la langue comme système de signes, axe syntagmatique / paradigmatique) - éléments sur la postérité du structuralisme : cercle de Prague, de Copenhague, fonctionnalisme, anthropologie structurale <p>Fonctions du langage</p> <ul style="list-style-type: none"> - schéma de communication (Roman Jakobson) <p>Analyse sémantique structurale et cognitive</p> <ul style="list-style-type: none"> - analyse sémantique structurale : l'analyse sémique , champs sémantiques - critiques de l'analyse sémique et théories du prototype

Introduction à la grammaire française : la phrase simple (18h CM, 2*18h TD, Licence Lettres Modernes L1S1)

Objectifs du cours	Etude de la phrase simple en français et de ses composants : morphèmes libres et liés, parties du discours, syntagmes et décomposition des syntagmes essentiels et circonstanciels.
Contenu du cours	Identification des constituants de la phrase simple et description de leurs propriétés grammaticales. Description des différentes catégories grammaticales des constituants (mot ou groupe de mots) et des différentes fonctions qu'ils peuvent occuper dans la phrase.

Introduction à la grammaire française : la phrase complexe (18h CM, 2*18h TD, Licence Lettres Modernes L1S2)

Objectifs du cours	Reprendre et approfondir les notions de la grammaire traditionnelle. Ce cours fait suite au cours « Introduction à la grammaire française : la phrase simple », traitant de la phrase simple et de ses composants.
Contenu du cours	<p>Rappel sur les notions essentielles en analyse grammaticale : classes de mots, fonctions, constituants de la proposition simple.</p> <p>Etude de la phrase complexe :</p> <ul style="list-style-type: none"> • Types de phrases : assertion, interrogation, exclamation, injonction, négation, emphase (dislocation et extraction), passif, impersonnel, phrases atypiques • Notion de proposition principale / proposition subordonnée • Les propositions relatives • Les propositions complétives • les propositions subordonnées circonstancielle • Juxtaposition et coordination

Problématiques de la néologie (24h CM, master 2 COLEDI)

Objectifs du cours	Ce séminaire a quatre objectifs principaux : sensibiliser les étudiants à l'importance de la néologie dans la vie des langues ; inculquer aux étudiants les notions théoriques utiles à l'analyse linguistique de la néologie lexicale ; inculquer aux étudiants les différentes approches pour repérer automatiquement et suivre le cycle de vie des néologismes, dans le cadre du traitement automatique des langues ; faire pratiquer les étudiants au repérage, à l'analyse et au suivi des néologismes, au travers de la plateforme Néoveille (www.neoveille.org). Un projet individuel permettant de valider le séminaire sera défini pour chaque étudiant.
Contenu du cours	Notions fondamentales : articulation langue/discours, synchronie/diachronie et dynamisme des langues Perspectives linguistique, sociolinguistique et cognitive sur le phénomène Notions d'unité lexicale et de construction Analyse linguistique des néologismes : catégorisation des néologismes, autres informations Application : détection et description d'innovations lexicales à partir de la plateforme néoveille.

Enseignements de linguistique informatique

Mes enseignements de linguistique informatique comprennent :

- des apprentissages techniques (programmation en PERL/PYTHON, Bases de données, Document numérique),
- des présentations générales du domaine, de ses applications et problématiques (introduction au TAL)
- des enseignements focalisant sur l'analyse automatique des textes au moyen de ressources linguistiques, de calculs statistiques ainsi que ceux utilisant des algorithmes d'apprentissage automatique. Ces derniers cours étant à la frontière avec mon travail de recherche, sont dispensés à partir du Master. Le cours « analyse automatique des textes en Python » donne lieu actuellement à la rédaction d'un manuel.

Bases de données et lexicographie (18h CM/TD, Licence lettres modernes et LLCE, L2/L3)

Objectifs du cours	Comprendre le fonctionnement d'une base de données dans la perspective de la gestion de dictionnaires et de corpus
Contenu du cours	La maîtrise des bases de données est une compétence et un savoir-faire incontournable en linguistique. Le cours constitue une initiation à l'issue de laquelle les étudiants seront capables de créer et modifier leurs propres bases de données, en adéquation avec leur projet professionnels et leurs intérêts propres, en illustrant le propos sur la lexicographie. Les différents éléments d'une base de données sont présentés au travers de deux logiciels : MS Access et MySQL, au travers de l'interface web phpmyadmin.

Document numérique : analyse structurée (18h CM/TD, Licence lettres modernes et LLCE, L2/L3)

Objectifs du cours	Il s'agit de présenter les bases de la structuration informatique des documents numériques. Pour sensibiliser les étudiants aux enjeux sous-jacents, on indiquera dans un premier temps les principales exploitations faites de la structuration des documents numériques et on présentera dans un deuxième temps les principes de modélisation et de programmation de cette structuration
Contenu du cours	Cet enseignement porte sur la structuration des documents numériques. <ul style="list-style-type: none"> • Historique du document numérique • Les encodages des documents numériques. • Les formats des documents numériques • Présentation détaillée du format XML • Normes TEI, OpenDocument, openXML • Niveaux de structuration des documents : méta-données, typologie des documents (noms TEI spécifiques), annotation des documents numériques

initiation au TAL (18h CM/TD, 3 ECTS, Licence lettres modernes et LLCE, L2/L3)

Objectifs du cours	Initier les étudiants au Traitement Automatique des Langues
Contenu du cours	Le cours présente les différents aspects de la linguistique informatique. Dans un premier temps, il est question du traitement informatique des principales caractéristiques des langues naturelles (analyse morphologique, analyse syntaxique, analyse sémantique). Dans un deuxième temps, les principales applications du Traitement Automatique des Langues (TAL) sont présentées, en décrivant et illustrant les applications issues de la compréhension et de la génération de textes (extraction d'information, recherche d'information, génération de textes). Le cours s'appuie sur la manipulation de logiciels de TAL existants : dictionnaires électroniques, Concordanciers, logiciels d'analyse linguistique

Introduction au TAL (24h CM, Master 1 TAL/COLEDI)

Objectifs du cours	Ce séminaire présente de façon détaillée les différentes opérations nécessaires pour effectuer une analyse automatique des textes au moyen de ressources linguistiques, et au moyen de calculs fréquentiels. Il présente également les techniques d'apprentissage automatique dites « supervisées », à partir de corpus annotés. Les présentations s'appuient sur des exemples linguistiques tirées de langues appartenant à des familles différentes, permettant ainsi d'appréhender les difficultés spécifiques à chaque langue.
Contenu du cours	<ul style="list-style-type: none"> • Récupération de corpus sur internet et pré-traitements • Segmentation des textes • Analyse morpho-syntaxique des textes • Analyse syntaxique des textes • Analyse sémantique des textes

Master 2 TILDE :

Les enseignements du Master TILDE sont professionnalisants. Leur objectif est de fournir des outils informatiques, une culture générale en TAL, et une pratique qui permettront aux étudiants d'aborder dans les meilleures conditions les problématiques réelles qu'ils rencontreront ensuite en entreprise.

Programmation en PERL (puis PYTHON depuis 2012) pour linguistes (30h CM, Master 2 TILDE)

Déroulement de l'enseignement	Cours présenciel sur 10 semaines de 3h. Support de cours disponible sur la plateforme d'enseignement à distance
Objectifs du cours	Maîtrise de la programmation en PERL/PYTHON permettant d'aborder les notions fondamentales de la programmation pour le langage PERL/PYTHON, très utilisé en TAL.
Contenu	<ul style="list-style-type: none"> • Prise en main de l'environnement de développement • Types de données • Boucles et structures de contrôle • Entrées / Sorties • Procédures et Modules • Debuggage des programmes • Références • Structures de données complexes • Programmation objet
Compétences et savoir-faire acquis à l'issue de la formation	<p>Compétences :</p> <ul style="list-style-type: none"> • programmation dans un langage structuré • résolution algorithmique et informatique de problèmes informatiques simples <p>Savoir-faire acquis :</p> <ul style="list-style-type: none"> • programmation en PERL/PYTHON basique (scripts utilitaires) • capacité à comprendre des programmes plus complexes • maîtrise des différentes composantes décrites dans le contenu.

Analyse automatique des textes (32h CM, Master 2 TILDE)

Objectifs du cours	Montrer aux étudiants les grandes étapes classiques de l'analyse automatique des textes au travers de programmes existants ou qui seront faits avec les étudiants (en Python). Exemples données dans différentes langues pour prendre conscience des traitements différenciés à appliquer.
Contenu	<ul style="list-style-type: none">• Encodage/Formats/conversion XML• calculs statistiques / probabilistes : état de l'art, applications à la détection d'unités polylexicales, word embeddings.• Récupération de données textuelles sur le web• Stockage des résultats dans un moteur de recherche (Apache Solr)• segmentation : état de l'art, application/programmation• analyse morphologique: état de l'art, application/programmation• analyse syntaxique: état de l'art, application/programmation• analyse sémantique: état de l'art, application/programmation

Autres projets de formation

Coordinateur Projet IDEX PSC « Pédagogies innovantes », 2015-2016 puis 2017-2018, Outillage numérique et informatique pour masters et doctorants en linguistique, littérature et langues au sein de SPC : formations intensives sur les outils et techniques informatiques utiles à la recherche dans ces domaines. 19 formations organisées en 2015-2016, 7 en 2017-2018 (<http://moodle.sorbonne-paris-cite.fr/course/view.php?id=239>, connexion anonyme)

Principales responsabilités administratives

Membre élu de la section 7 du Comité National des Universités : 2011-2016

Co-responsable de la formation du Labex *Empirical Foundations of Linguistics* (EFL) depuis 2015, puis responsable depuis 2018

Directeur du département de linguistique de l'université Paris 13 (12 enseignants-chercheurs, de 2 à 4 ATER, un secrétaire pédagogique) : 2010-2014

Membre élu du Conseil d'Administration de l'Université Paris 13 Sorbonne Paris Cité : 2016-2018

Membre élu du conseil d'UFR LLSHS (Université Paris 13) depuis décembre 2018

Logiciels et ressources linguistiques

Néoveille, plateforme de détection, de description et de suivi des néologismes en corpus dynamique : <http://www.neoveille.org>. Un échantillon de 13 000 néologismes et informations associées est disponible sur la partie publique (menu Néologismes récents à gauche). Pour accéder aux données complètes, cliquer sur l'item de menu (à gauche) connexion (identifiant et mot de passe : neoveille-demo). On peut consulter les corpus (menu Corpus/Fils RSS-Sites web), les néologismes candidats (menu Néologismes de forme/Néologismes candidats), et les néologismes validés (menu gestionnaire des néologismes/base générale(srv)).

Néonaute, moteur de recherche pour suivre l'implantation des néologismes à partir des collections du dépôt légal du web (BnF) et de Néoveille : <https://tal.lipn.univ-paris13.fr/neonaute/html/>. Pour accéder aux données, cliquer sur l'item de menu (à gauche) connexion (identifiant et mot de passe : neonaute-demo). Les listes de termes étudiés sont accessibles par le menu « listes » à gauche.

Pour les autres logiciels et ressources, on consultera la synthèse du dossier d'HDR (conclusion de chaque chapitre).

Publications

On trouvera une version numérique de la majorité des publications à cette adresse : <https://lipn.univ-paris13.fr/neoveille/html/data/ecartier/publications/>

Le dossier d'HDR complet se trouve à cette adresse : <https://lipn.univ-paris13.fr/neoveille/html/data/ecartier/> (CV étendu, Document de synthèse, Inédit, liste des principales publications par thématiques)

Ouvrages

Cartier E. (à paraître, 2019). « Dynamique lexicale des langues : éléments théoriques, méthodes automatiques , expérimentations en français contemporain », document inédit HDR, 237 pages.

Chapitre d'ouvrages

Cartier E. et Boutmgharine N. (à paraître, 2019). « L'innovation lexicale en français : une étude comparative de la presse féminine et de la presse généraliste », Dans Hadermann P., Dostie G., Lefeuvre F et Diermoz F. (eds), *Le français innovant, Collection Sciences pour la communication, Peter Lang*.

Cartier E. (2018). « Emprunts en français contemporain : étude linguistique et statistique à partir de la plateforme Néoveille » dans *Emprunts en question(s)*, Kacprzak, A. ; Mudrochová, R. ; Sablayrolles, J.-F. (éds), La Lexicothèque, Limoges, Lambert-Lucas, 27p.

Cartier E. (2016), "Distributionnalisme et sémantique : état des lieux en traitement automatique des langues", In MARTINOT, Claire, MARQUE-PUCHEU, Christiane, GEROLIMICH, Sonia, (éds), 2016, *Perspectives harrisiennes*, Paris: CRL, 344p. , pp.288-313.

Cartier E. (2008), "Intégration des prédicats verbaux dans l'analyseur sémantique TextBox : l'exemple des verbes de cognition", dans *Verbum « Les prédicats verbaux »*, p. 97-111, 2008.

Cartier E. (2008). « Repérage automatique des expressions figées : état des lieux, perspectives ». In Peter Blumenthal ; Salah Mejri (ed.), *Les séquences figées : entre langue et discours, Zeitschrift für Französische Sprache und Literatur*, pages 55-70. Franz Steiner Verlag.

Cartier E. (2000). « Éléments pour une modélisation des dictionnaires électroniques ». In *Dictionnaires bilingues, Méthodes et contenus*, sous la direction de Thomas Szende,

Paris, H. Champion, pages 135–152.

Articles dans des revues à comité de lecture

- Cartier E. (à paraître 2019), « Neoveille, système de repérage et de suivi des néologismes en onze langues », *Néologica*.
- Cartier E. (2018), « Noms propres et innovations lexicales : étude linguistique et statistique à partir de Néoveille », *Cahiers de Lexicologie*, 18p.
- Cartier E. (2016). "Neoveille, système de repérage et de suivi des néologismes en sept langues", *Neologica* 10, p. 101-131.
- Cartier E. (2012). "Néologie et Traitement Automatique des Langues", Numéro Spécial « la néologie », *Langages*.
- Cartier E. (2012). « Dicionários semânticos do LDI: modelo teórico, ferramenta de gestão computacional, ferramenta de exploração ». In: Aparecida Negri Isquerdo, Maria Candida Trindade Costa de Seabra (orgs.). *As Ciências do Léxico: lexicologia, lexicografia, terminologia*. vol. 6. Campo Grande: Editora UFMS, 2012. p. 261-282. 539p.
- Cartier E. (2011). "Étiquetage sémantique des textes : état des lieux, éléments de modélisation", Numéro Spécial « Catégories linguistiques et étiquetage de corpus », *L'information grammaticale*, juin 2009.
- Cartier E. (2011). « Néologie et description linguistique pour le TAL ». *Langages*, vol. 183, no.3. pp.105-117
- Cartier E. (2009). « Dictionnaires sémantiques électroniques : exemple de l'implémentation du modèle du LDI ». *Etudes de linguistique appliquée*, (4), p. 76-89.
- Cartier E. (2009). « Étiquetage, sémantique des textes: états des lieux, éléments de modélisation ». *L'Information Grammaticale*, 122(1):19–29.
- Cartier E. and Issac F. (2009). « Mot et traitement automatique des langues ». *Le Français Moderne - Revue de linguistique Française*, 77(1):145–160.
- Cartier E. and Sablayrolles J.-F. (2009). « Néologismes, dictionnaires et informatique ». *Cahiers de Lexicologie*, 2008-2(93):175–192.
- Cartier E. et Sablayrolles J.-F. (2008). "Néologismes, Dictionnaires et Informatique", *Les Cahiers de Lexicologie*, n° 93, 2008- 2, p. 175-192.
- Minel J.-L., Desclès J.-P., Cartier E., Crispino G., Ben Hazez S., Jackiewicz A. (2001). « Résumé automatique par filtrage sémantique d'informations dans des textes. Présentation de la plate-forme FilText », *Revue Technique et Science Informatique*, n° 3, 2001.
- Berri J., Cartier E., Desclès J.-P., Jackiewicz A., and Minel J.-L. (1996). « A linguistic method for text filtering ». *Revista del Procesamiento del lenguaje natural*, 19:159–165.

Communications publiées avec comité de lecture – conférences internationales

- Cartier E., Sablayrolles J.-F., Boutmgharine N., Humbley J., Bertocci M., Jacquet-Pfau C., Kübler N. et Tallarico G. (2018), « Détection automatique, description linguistique et suivi des néologismes en corpus : point d'étape sur les tendances du français contemporain », *Actes du Congrès Mondial de Linguistique Française*, Mons (Belgique), 9-13 juillet 2018, 20p.
- Cartier E. (2017), « Neoveille, a Web Platform for Neologism Tracking », *European Chapter of the Association for Computational Linguistics* 22-27 Avril 2017.

- Cartier E. (2017), « Detection and Life-cycles Tracking of Neologisms : experiments in the Neoveille Project », *Proceedings of The Dynamics of Lexical Innovation*, Munich, 28-30 juin 2017.
- Cartier E., Gabor K., Lejeune G., et Charnois T. (2017). « Néoveille, a Web Platform for Neologism Tracking ». In *Supplementary Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2017)*, pages 13–25.
- Viaux J., Cartier E. (2016), « Étude linguistique et quantitative de la pénétration des anglicismes de type (N,ADJ)-Ving dans sept langues à partir d'un corpus contemporain journalistique », *Colloque international Emprunts néologiques et équivalents autochtones. Mesure de leurs circulations respectives*, University of Łódź, 10-12 octobre 2016, Łódź, Pologne.
- Cartier E. (2015). « Linguistic preprocessing for distributional analysis : Evidence from French ». In *Corpus Linguistics 2015*, Lancaster, United Kingdom. Lancaster University.
- Cartier E. and Gonzalez P. C. (2012). « Technological Tools for Dictionary and Corpora Building for Minority Languages: example of the French-based Creoles ». *Workshop Language Technology for Normalisation of Less- Resourced Languages, LREC 2012*, pages 47–59.
- Cartier E. (2011). « Repérage automatique des séquences figées : état des lieux, perspectives ». In *Proceedings of IIIe Jornadas internacionales de fraseologia contrastiva : fraseologia opacidad y traducción*, 18-19 novembre 2010, Université d'Alicante, Espagne.
- Cartier E. (2011). « Utilisation des contextes dans le cadre dictionnaire : état des lieux, typologie des contextes, exemple des contextes définitoires ». In *Actes des Huitièmes Journées scientifiques du Réseau de chercheurs Lexicologie, terminologie, traduction*, Lisbonne, 15-17 octobre 2009. Editions des Archives Contemporaines et Agence universitaire de la Francophonie, pages 619–632.
- Joseph A. et Cartier E. (2011). « Repérage automatique des séquences figées pour la classification des documents ». In *Proceedings of LTT 2011*.
- Cartier E. (2009). « Corpus for linguistic resources building and maintenance (CLRBM): Architecture, first experiments ». In *Proceedings of Corpus Linguistics 2009*, Lancaster, United Kingdom. Lancaster University.
- Cartier, E. and Sablayrolles, J.-F. (2008). « Neologia, une base de donnée pour la gestion des néologismes. » In Cabré, Maria Teresa and Domènech Bagaria, Ona and Estopa, Rosa and Lorente, Mercè, editor, *Congrés International de Neologia dans les Langues romanes*, pp.759–767, Barcelone, Spain. IULA.
- Cartier E. (2007). « TextBox, a Written Corpus Tool for Linguistic Analysis ». In Cédric Fairon ; Hubert Naets ; Adam Kilgarrieff ; Gilles-Maurice De Schryver, editor, *Web As Corpus 2007, Cahiers du CENTAL*, pages 33–42, Louvain-la-Neuve, Belgium. Presses universitaires de Louvain.
- Hachey B., Grover C., Karkaletsis V., Valarakos A., Paziienza M. T., Vindigni M., Cartier E., Coch J., et al. (2003). « Use of ontologies for cross-lingual information management in the web ». In *Proceedings of the Ontologies and Information Extraction International Workshop held as part of the EUROLAN 2003, Romania, July 28-August 8 2003*.
- Karkaletsis V., Spyropoulos C. D., Souflis D., Grover C., Hachey B., Paziienza M. T., Vindigni M., Cartier E., et Coch J. (2003). « Demonstration of the CROSSMARC system ». In Hearst, Marti A. and Ostendorf, Mari, editor, *Proceedings HLT-NAACL 2003, ACL*.
- Lenci A., Bartolini R., Calzolari N., Agua A., Busemann S., Cartier E., Chevreau K., and Coch J. (2002). « Multilingual summarization by integrating linguistic resources in the mlis-musi project ». In *Proceedings of LREC 2002*, volume 2, pages 1464–1471

- Desclés J.-P., Cartier E., Jackiewicz A., et Minel J.-L. (1997). « Textual processing and contextual exploration method ». *Proceedings of Context 97*, Universidade Federal do Rio de Janeiro, Brésil, 97:189–197.
- Berri J., Cartier E., Desclés J., Jackiewicz A., et Minel J. (1996). « Filtrage automatique de textes ». *Natural Language Processing and Industrial Applications*, pages 28–35, Moncton, Canada.

Communications publiées avec comité de lecture – conférences nationales

- Cartier E. (2015). « Extraction automatique de relations sémantiques dans les définitions : approche hybride, construction d'un corpus de relations sémantiques pour le français ». In *Conférence annuelle Traitement Automatique des Langues Naturelles*, Caen, France.
- Cartier E. (2015). "Extraction automatique de relations sémantiques dans les définitions : approche hybride, construction d'un corpus de relations sémantiques pour le français", *TALN 2015*, Caen, 22-25 juin 2015.
- Grezka A., Cartier E., et Mathieu-Colas M. (2015). « Dictionnaires morphologiques du français contemporain : présentation de morfetik, éléments d'un modèle pour le TAL ». In *Traitement Automatique des Langues Naturelles (TALN)*, Caen, France. Université de Caen.
- Cartier E. et Sablayrolles J.-F. c. (2009). « Nouvelles technologies, nouveaux modèles linguistiques et néologie ». In *Sciences du langage et nouvelles technologies*, colloque 2009 de l'Association des Sciences du Langage, pages 53–59, Paris, France. Lambert-Lucas.
- Buvet P.-A., Cartier E., Issac F., Mathieu-Colas M., Mejri S., et Madiouni Y. (2009). « Morfetik, ressource lexicale pour le TAL ». In *Proceedings of TALN 2009*, Senlis, 24-26 juin 2009, pages 1–10.
- Buvet P.-A., Cartier E., Issac F. et Mejri S. (2007). « Dictionnaires électroniques et étiquetage syntactico-sémantique ». In Hathout, Nabil ; Muller, Philippe, editor, *Proceedings of TALN 2007*, pages 239–248, Toulouse, France. IRIT Press.
- Cartier E. (2005). "Repérage automatique des informations définitives dans e-doc Explorer", *TIA 2005*, 4-5 avril 2005.
- Minel J.-L., Desclés J.-P., Cartier E., Crispino G., Ben Hazez S., et Jackiewicz A. (2001). « Résumé automatique par filtrage sémantique d'informations dans des textes ». *Traitement automatique du langage naturel*, 20(3):369–395.
- Cartier E. (1998). « Analyse automatique des textes: l'exemple des informations définitives ». In *Actes Rencontre Internationale sur l'Extraction, le Filtrage et le Résumé Automatiques*, pages 6–18.
- Cartier E. (1997). « La définition dans les textes scientifiques et techniques: présentation d'un outil d'extraction automatique des relations définitives ». *Proceedings of TIA 97*, pages 41–48.
- Berri J., Cartier E., Desclés J.-P., Jackiewicz A., et Minel J.-L. (1996). « Safir, système automatique de filtrage de textes ». In *Proceedings of the TALN Annual Conference*, Marseille, 22-24 mai 1996.

Communications non formellement publiées

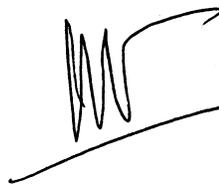
- Cartier E., Stirling P. et Aubry S. (2018): « Néonaute: mining web archives for linguistic analysis », in *IIPC Web archiving Conference 2018 Proceedings*, Wellington (New-Zealand), 12-15 nov. 2018.

- Cartier E. (2017), « Sémantique lexicale et distributionnalisme : éléments pour le repérage automatique du sens en corpus », *Proceedings of the 7th Représentations du sens linguistique*, Sherbrooke, 25-27 oct. 2017.
- Cartier E. (2016). "Word Profile Automatic Tracking Through the Prism of Construction Grammars and Distributionalism", 9th International Conference on Construction Grammar (ICCG9), Federal University of Juiz de Fora, Brazil, October 5th-7th 2016.
- Cartier E. (2016). « Word Profile Automatic Tracking Through the Prism of Construction Grammars and Distributionalism ». In *9th International Conference on Construction Grammar (ICCG9)*, Juiz de Fora, Brazil.
- Cartier E. et Vinogradova V. (2015). "Linguistic Preprocessing for Distributional Analysis : Evidence from French", *Corpus Linguistics 2015*, Lancaster, 17-21 july 2015.
- Cartier E. (2010). « Inférence et traitement automatique des langues : état des lieux, modélisation, prototype ». In *Colloque international Gerflint 3-5 juin 2010*, Cracovie.
- Cartier E. et Grezka A. (2009). "Syntactico-semantic Classification of Verbs : a case-study", *Gent*, 5-7 février 2009.
- Cartier E., Chevreau K., et Coch J. (2002). « Résumé orienté par la question et filtrage ». In *Journée d'étude ATALA : la résumé automatique de texte : solutions et perspectives*, 14 déc. 2002, Paris, France.
- Cartier E. (1998). *Analyse Automatique des textes: L'exemple des expressions définitoires*, RIFRA 98, Sfax. 11-14 Novembre 1998.
- Cartier E., Desclés J-P., Jackiewicz A. Minel J.-L., (1997). *Filtrage automatique de textes : l'exemple des énoncés définitoires " 1° JST FRANCIL 97*, Avignon, 1997.

Rapports techniques

- Cartier E. (2006). « e-doc anonymizer, une application pour l'anonymisation des noms de personnes dans les arrêts de la cour des comptes ». *Technical report, e-doc Labs*.
- Cartier E. (2006). « e-doc wineexplorer, reconnaissance automatique multilingue des vins vendus en ligne ». *Technical report, e-doc Labs*.
- Lenci A., Bartolini R., Calzolari N., Agua A., Busemann S., Cartier E., Chevreau K., et Coch J. (2001). « Architecture specifications, deliverable d1.1, workpackage 1 », MLIS-MUSI (European Project MLIS-5015). *Technical report*.
- Lenci A., Bartolini R., Calzolari N., et Cartier E. (2001). « Document analysis, MLIS-MUSI (European Project MLIS-5015), deliverable d3.1, workpackage 3 ». *Technical report*.

Paris, le 17/12/2018



Emmanuel
CARTIER