

Nicolas Hernandez

1. Identification

- Université de Nantes / IUT de Nantes (depuis 2006)
- Laboratoire des Sciences du Numérique de Nantes (LS2N UMR CNRS 6004)

2. Activités exercées durant les 4 dernières années

Mon domaine de recherche se trouve à la jonction des domaines du Traitement Automatique des Langues (TAL) et de la Fouille de Données Textuelles. Ma thématique gravite autour des questions de modélisation linguistique, numérique et d'analyse automatique de l'organisation de l'information au niveau discursif. Depuis 2014, je m'intéresse plus particulièrement aux traitements des Communications écrites Médiées par les Réseaux (CMR e.g. forum, chat, mails) et, depuis 2016, au domaine applicatif de l'enseignement en ligne.

2.1 Publications et production scientifique

Les publications ci-dessous couvrent mon activité de **2017 à 2020**. Elles se répartissent sur 4 thèmes :

- Modélisation, notamment “profonde”, d'information discursive au sein des CMR (actes du dialogue portés par un énoncé, relation rhétorique et dialogique entre énoncés consécutifs) [3, 5, 6, 14, 15, 16] avec 2 doctorants, Soufian Salim et Basma Boussaha
- Modélisation profonde et calcul de similarité, voire d'alignement, de contenus textuels [8, 9, 10, 12] avec 1 post-doctorant, Amir Hazem
- Reconnaissance de la parole dans un contexte de cours magistraux [4, 7, 13] avec 1 doctorant, Salima Mdhaffar
- Production de ressources langagières et logicielles pour la fouille de données textuelles [7, 11] avec Salima Mdhaffar et Soufian Salim

Le **rang** des conférences suit la classification *Computing Research and Education Association of Australasia*¹ (**CORE**). Le tableau ci-dessous en fournit une synthèse.

Activité éditoriale	2
Revue internationale à comité de lecture (CORE A)	1
Conférences internationales (CORE A)	1

¹ <http://www.core.edu.au/conference-portal>

Conférences internationales (CORE B)	2
Conférences internationales (CORE C)	6
Conférences nationales avec comité de sélection et actes (CORE C)	2
Ateliers internationaux (avec comité de sélection et actes)	2
TOTAL	16

Activité éditoriale

1. Nicolas Hernandez et Lydia Mai Ho-Dac, Edition du numéro Varia (26), “Discours, revue électronique de linguistique, psycholinguistique et informatique”, Presse Universitaire de Caen, Juin 2020
2. Nicolas Hernandez et Christian Raymond, Edition du numéro Varia (20), “Discours, revue électronique de linguistique, psycholinguistique et informatique”, Presse Universitaire de Caen, Déc 2017

Articles dans revues internationales à comité de lecture (CORE A)

3. Basma El Amel Boussaha, Nicolas Hernandez, Christine Jacquin, Emmanuel Morin. End-to-End Response Selection Based on Multi-Level Context Response Matching, *Computer Speech & Language*, Elsevier, 2020

Conférences internationales (CORE A)

4. Salima Mdhaffar, Yannick Estève, Nicolas Hernandez, Antoine Laurent, Richard Dufour, Solen Quiniou. Qualitative Evaluation of ASR Adaptation in a Lecture Context: Application to the PASTEL Corpus, *Interspeech*, session poster, Graz, Austria. pp.569-573, Sep 2019

Conférences internationales (CORE B)

5. Basma El Amel Boussaha, Nicolas Hernandez, Christine Jacquin, Emmanuel Morin. Towards Simple but Efficient Next Utterance Ranking, In *20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, session poster, La Rochelle, France, Avr 2019
6. Soufian Salim, Nicolas Hernandez, and Emmanuel Morin. Dialogue Act Taxonomy Interoperability Using a Meta-Model. In *18th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, article long, Budapest, Hungary, Avr 2017

Conférences internationales (CORE C)

7. Salima Mdhaffar, Yannick Estève, Antoine Laurent, Nicolas Hernandez, Richard Dufour, Delphine Charlet, Geraldine Damnati, Solen Quiniou, Nathalie Camelin. A Multimodal Educational Corpus of Oral Courses: Annotation, Analysis and Case Study. In

Proceedings of the 12th Language Resources and Evaluation Conference (LREC), article long, 2020

8. Amir Hazem, Nicolas Hernandez. Tweaks and Tricks for Word Embedding Disruptions, In *Recent Advances In Natural Language Processing (RANLP)*, session poster, Varna, Bulgaria, Sep 2019
9. Amir Hazem, Nicolas Hernandez. Meta-Embedding Sentence Representation for Textual Similarity, In *Recent Advances In Natural Language Processing (RANLP)*, session poster, Varna, Bulgaria, Sep 2019
10. Amir Hazem, Nicolas Hernandez, Basma Boussaha. A multi-domain framework for textual similarity. a case study on question-to-question and question-answering similarity tasks. In *11th edition of the Language Resources and Evaluation Conference (LREC)*, session poster, Miyazaki, Japon, 7-12 Mai 2018
11. Nicolas Hernandez, Amir Hazem. Pyrata, python rule-based feature structure analysis. In *11th edition of the Language Resources and Evaluation Conference (LREC)*, session poster, Miyazaki, Japon, 7-12 Mai 2018
12. Amir Hazem, Basma El Amel Boussaha, Nicolas Hernandez. Mappsent: a textual mapping approach for question-to-question similarity. In *Recent Advances in Natural Language Processing (RANLP)*, article long, Varna, Bulgarie, Sep 2017

Conférences nationales avec comité de sélection et actes (CORE C)

13. Salima Mdhaffar, Yannick Estève, Nicolas Hernandez, Antoine Laurent, Solen Quiniou. Apport de l'adaptation automatique des modèles de langage pour la reconnaissance de la parole : évaluation qualitative extrinsèque dans un contexte de traitement de cours magistraux, In *26e conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, article court, Toulouse, France, 2019
14. Basma El Amel Boussaha, Nicolas Hernandez, Christine Jacquin, Emmanuel Morin. Ordonnancement de réponses dans les systèmes de dialogue basé sur une similarité contexte/réponse. In *25e conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, article long, Rennes, France, Mai 2018

Ateliers internationaux (avec comité de sélection et actes)

15. Basma El Amel Boussaha, Nicolas Hernandez, Christine Jacquin, Emmanuel Morin. Multi-level Context Response Matching in *Retrieval-Based Dialog Systems Dialog System Technology Challenges Workshop (DSTC 7) at AAAI*, Honolulu, United States, Jan 2019
16. Basma El Amel Boussaha, Nicolas Hernandez, Christine Jacquin, Emmanuel Morin. Next utterance ranking based on context response similarity, *5th Machine Learning and Data Analytics Symposium (MLDAS)*, Doha, Qatar, Mar 2018

2.2 Encadrement doctoral et scientifique

Sur la période, j'ai participé à l'encadrement de

- **3 doctorants** (1 financement FUI, 1 MESR, et 1 ANR), **tous avec leur thèse soutenue**, dans le domaine de l'Informatique et du Traitement Automatique des Langues
- **1 post doctorant**
- **1 Master 2**
- **4 stagiaires DUT 2, Licence 3 et Master 1**

Encadrement doctoral (thèse soutenues)

Soufian SALIM	01/10/2014- 30/11/2017 (38 mois)	Financement : FUI ODISAE "Analyse en actes du dialogues des conversations écrites en ligne orientées par la tâche" Directeur de thèse : E. Morin 40% (LINA) ; Encadrant : <u>60%</u> ; Publications : 1 CORE B, 2 CORE C Devenir : Chercheur chez Bee4win
Basma El Amel BOUSSAHA	01/10/2016 - 30/11/2019 (38 mois)	Financement : MESR "Response Selection for End-to-End Retrieval-Based Dialogue Systems" Directeur de thèse : E. Morin 40% (LS2N) ; Encadrants : <u>N. Hernandez 30%</u> et C. Jacquin 30% (LS2N) Publications : 1 CORE A, 1 CORE B, 3 CORE C Devenir : Senior data scientist/NLP chez della.ia (startup)
Salima MDHAFAR	01/01/2017 - 30/04/2020 (40 mois)	Financement : ANR PASTEL "Reconnaissance de la parole dans un contexte de cours magistraux : évaluation, avancées et enrichissement" Directeur de thèse : Y. Estève 25% (LIUM) ; Encadrants : <u>N. Hernandez 25%</u> ; A. Laurent 25% (LIUM), S. Quiniou 25% (LS2N) Publications : 1 CORE A, 1 CORE C Devenir : post-doctorat

Encadrement post-doctoral

Amir HAZEM	Juillet 2017 (11 mois)	Financement : ANR PASTEL "Modélisation profonde du sens des mots et de passages textuelles et alignement multimodal" - https://github.com/hazemAmir/MappSent Publications : 5 CORE C Devenir : post-doctorat au LS2N
------------	------------------------	---

Encadrement Master 2

Samy GASCOIN-FONTAINE	Octobre 2020 (3 mois)	Financement : ANR PASTEL "Valorisation - paquetage de données collectées et produites durant le projet ANR PASTEL"
Esther FELIX	Février 2021 (6 mois)	Financement : LS2N

“Apport du TAL pour l’assistance à l’apprentissage humain - mesures de l’engagement des apprenants dans les forums de MOOCs”
Co-encadrement à 80% avec Issam Rebaï à 20 % (IMT Atlantique)

Autres encadrements de stagiaires (DUT, Licence et Master 1)

Yanis OUAKRIM	Avril 2017 (3 mois) DUT Info 2e année Financement : ANR PASTEL “Développement d’une bibliothèque d’extraction d’information au sein de diaporamas multi-formats PPTX, ODP et Latex” - https://github.com/Codophile1/exide
Matthias GOULAY	Avril 2018 (2,5 mois) Licence Info 3 Financement : ANR PASTEL “Développement et expérimentation d’un extracteur de mots clefs au sein de transcriptions de discours d’enseignants en cours magistraux et des diaporamas associés” - https://github.com/MGoulley/PySTKE
Axel RAMBAUD	Avril 2019 (2,5 mois) Master Info 1 Financement : ANR PASTEL “Expérimentation d’approches état de l’art pour aligner la transcription d’un discours d’un enseignant en cours magistral avec le diaporama associé”
Maëlle BRASSIER	Avril 2019 (2,5 mois) Master Info 1 Financement : ANR PASTEL “Définition et expérimentation de mesure automatique de la notion de difficulté et de concepts difficiles dans un cours”

2.3 Diffusion de travaux (rayonnement et vulgarisation)

Expertise (organismes nationaux ou internationaux)

Je participe à la vie de la communauté francophone de mon domaine de recherche à travers des relectures pour la revue et les conférences associées. Aujourd’hui mon implication dans les organes d’évaluation et de financement nationaux se traduit à travers l’expertise de soumission de projets à l’ANR et de sujets de thèse CIFRE.

Ne sont listées que les expertises de la période :

- 1 Comité de lecture de revues
2018 Traitement Automatique des Langues (TAL) (revue internationale classée au rang A par l’AERES en SHS)
- 6 Comités de lectures de conférences
2017, 2018, 2019, 2020 Traitement Automatique des Langues Naturelles (TALN)
Congrès francophone annuel du domaine
2019 Rencontres des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RECITAL)
Congrès annuel pour les jeunes chercheurs du domaine

2017 ACS/IEEE International Conference on Computer Systems and Applications

- 1 projet (régions, ANR, Europe, AERES, ...)

2019 Appel Générique de l'ANR : expertise de la qualité scientifique, de l'organisation et de l'implémentation du projet, de son impact économique, social et culturel ainsi que de la pertinence et complémentarité des partenaires

- 2 thèses CIFRE

2018, 2019 Pour l'Association nationale de la recherche et de la technologie (ANRT) : Expertise de la qualité scientifique du projet, de la pertinence du contexte entrepreneurial, de l'adéquation du laboratoire, et du profil du candidat

Activités éditoriales (expertises, responsabilités de collections...)

Juin 2020 *Co-éditeur* du numéro varia 26 de la revue électronique Discours Appel, recherche d'experts, suivi des articles, suivi des experts, traduction et publication ; hors appel, 6 mois de travail avec activité d'1-2h hebdo

Déc 2017 *Co-éditeur* du numéro varia 20 de la revue électronique Discours Appel, recherche d'experts, suivi des articles, suivi des experts, traduction et publication ; hors appel, 6 mois de travail avec activité d'1-2h hebdo

Organisation colloques, conférences, journées d'étude

2019 (7/10 et 4/11) Organisation de deux micro-ateliers pour la restitution des travaux réalisés dans le cadre du projet structurant CominLabs AILE² (AI for Learning Environments)

Ressources logicielles et langagières (développement, diffusion et maintenance)

*PyRATA*³ **Bibliothèque logicielle** "Python Rule-based feAture sTructure Analysis"
Module python 3 pour exprimer des motifs de recherche sur des annotations sémantiques ; licence Apache-v2
1ère diffusion : mars 2017 Au 24/02/2021 : 55 "stars" et 17 "fork"

RAD **Logiciel** de Reconnaissance en Actes du Dialogue⁴ pour les communications écrites médiées par les réseaux en français ; production faisant suite au projet porté FUI ODISAE terminé 1 an auparavant ; licence Apache-v2.
1ère diffusion : Mai 2017

² <http://aile.comin-ocw.org/news.html>

³ <https://github.com/nicolashernandez/PyRATA>

⁴ <https://goo.gl/forms/QDfs72kTSYkfGUt82>

Information sur le métier de chercheur et la thématique de recherche qu'est le Traitement Automatique du Langage Naturel

- Chaque année, participation à un **module d'initiation à la recherche** à l'UFR des Sciences en proposant un sujet de recherche et guidant dans la démarche scientifique un groupe de 3 étudiants de M1 durant 16 semaines (qui travaillent sur leur projet à raison d'une journée par semaine) ; point hebdomadaire d'1h
- Depuis 2019, participation aux "Rencontres de la recherche IUT" sur ½ journée par an échange d'1h avec des groupes d'une trentaine d'étudiants de 1ère année (toute discipline) pour **présenter ce qu'est la recherche, le métier de chercheur, le parcours pour y arriver et une thématique d'un travail de recherche**

2.4 Responsabilités scientifiques

Activités éditoriales

Depuis 2007 Membre du comité de rédaction de la revue électronique internationale et multidisciplinaire (linguistique, psycholinguistique et informatique) **Discours**⁵, avec relecture en double-aveugle qui publie 2 numéros par an en accès ouvert. *Discours* est soutenue par l'INSHS du CNRS et est répertoriée dans les bases telles que SCOPUS, DOAJ et ERIH PLUS. Le taux d'acceptation est de 25 %.

2-5 jours par an

Participation à la définition de la politique scientifique et éditoriale ; soutien aux coordinateurs et rédacteurs en chefs ; conseil et expertise scientifique

Depuis 2019 Co-rédacteur en chef de la revue Discours

1 à 2h toutes les deux semaines.

La mission concerne plusieurs tâches : recherche de financement, communication de la revue, gestion des aspects légaux, supervision des éditions des numéros, capitalisation sur le travail éditorial de chaque éditeur, gestion du comité scientifique de la revue, animation du comité de rédaction.

Montage et direction de projets : 2 projets co-montés financés

2019 Labex CominLabs "projet exploratoire" AILE⁶ (Artificial Intelligence for Learning Environments) visant à rassembler des chercheurs et consolider l'écosystème régional autour des questions de l'intelligence artificielle et de l'éducation
Montant alloué 5 675 €

Porteur avec Olivier Aubert ; partenaires : LS2N (équipes Duke et TALN), IMT Atlantique, Centre de développement pédagogique de l'Univ. de Nantes
Actions : organisation de 2 ½ journées de séminaires thématiques, financement de stages, organisation d'un atelier national, recrutement et direction de

⁵ <https://journals.openedition.org/discours/>

⁶ <http://aile.comin-ocw.org/>

stagiaires, exposés des travaux réalisés, foyer de montage d'au moins deux projets de recherche multipartenaires

2016-2021 **ANR/PRCE PASTEL**⁷ (Transcription Automatique de la Parole pour l'Apprentissage et la Formation) ; 42+12 mois ; subvention 98 118€ / 686 242 €
Copilotage, coordinateur pour le LS2N, responsable d'un lot technique ; partenaires : LIUM (porteur), CREN, Orange Lab, LS2N
Actions : Co-pilotage avec une réunion mensuelle de coordination avec partenaires et responsable local ; co-direction d'un doctorant, direction d'un post-doctorant et de stagiaires ; mise en place de moteurs d'enrichissement de contenu à partir de diaporama et de discours d'un enseignant en cours magistral, mesure de la difficulté de concepts, recherche d'information contrôlée et alignement de ressources pédagogiques

Projets avec d'autres formes de participation

2017-2020 **Projet européen Horizon 2020 X5-GON**⁸ (Cross Modal, Cross Cultural, Cross Lingual, Cross Domain, and Cross Site Global Open Educational Resources Network)
Réfèrent TALN dans l'équipe locale au LS2N ; essentiellement une activité de conseil et de réflexion sur la problématique d'indexation et de recommandation de ressources éducatives libres (REL), de modélisations profondes des contenus textuels et de mesure de la difficulté d'un concept, d'une ressource ou d'un passage d'une ressource (éventuellement une ressource vidéo à travers sa transcription)

Responsabilités administratives scientifiques

Depuis 2006 Réfèrent de l'équipe LS2N TALN sur les questions d'infrastructure logicielle et matérielle de l'équipe

Par le passé

- *Porteur d'un projet d'unification des développements de l'équipe* autour de la plateforme de développement de composants d'analyse de contenu Apache UIMA (Unstructured Information Management Architecture) ; Termsuite⁹, livrable du projet EU TTC, repose sur cette architecture
- *A l'initiative de l'outillage en logiciels collaboratifs et formation de l'équipe* (wiki, système de gestion de versions...)

Aujourd'hui

- *Soutien à l'activité deep learning* : analyse des besoins de l'équipe, identification du matériel requis, devis, achat et supervision de

⁷ <https://projets-lium.univ-lemans.fr/pastel>

⁸ <https://www.x5gon.org>

⁹ <http://termsuite.github.io/>

2.5 Autres activités et responsabilités (pédagogiques, administratives)

Rattaché au département Informatique de l'IUT de Nantes, l'équipe pédagogique **est extrêmement sollicitée et le département demande beaucoup d'investissement** (par exemple 2 années sur 3 sont proposées en sus en alternance). En moyenne, en plus de mes 192 h de service statutaire annuel, je cumule l'équivalent d'environ **148h (+ de 75% d'un service) heures complémentaires (réparties approximativement en 1/3 d'enseignements supplémentaires, 1/3 de responsabilités pédagogiques et 1/3 d'encadrement projets/stages)**.

Responsabilités administratives pédagogiques

Depuis sept 2013 *Responsable des stages pour le département INFO de l'IUT de Nantes*
Ma mission principale est d'accompagner **nos 88 étudiants de DUT2 et nos 20 étudiants de Licence Pro** dans la recherche et la réalisation d'un stage en entreprise.

Le travail me prend en moyenne **1 à 2 h hebdomadaire sur l'année** avec une sollicitation accrue les quelques mois qui précèdent le départ en stage (janvier-avril). J'ai la chance de bénéficier d'un soutien logistique de la part de mon secrétariat. Mes missions tournent autour de 3 activités :

1. **Relation avec les entreprises** : prospection, mailing pour solliciter des offres, information sur nos formations, sur des candidats éventuels, organisation de rencontres, diffusion des offres de stage/d'emploi
2. **Accompagnement des étudiants** : réunion d'information, simulation de candidatures électroniques, conseils sur leur candidatures, conseil pour chercher/trouver un stage, diffusion d'offres, validation individuelle de chaque mission, conseil dans le déroulement du stage, conseil pour la soutenance et écriture du rapport final, organisation des soutenances, relais avec le service international pour les stages à l'étranger, référent pour les étudiants entrepreneur...
3. **Supervision du suivi des enseignants tuteurs** : désignation, consignes, organisation des soutenances

J'ai produit et je maintiens un guide d'accompagnement pour la recherche et la conduite de stage à l'attention des étudiants de DUT et en Licence Professionnelle en IUT d'Informatique qui peut être consulté librement en ligne¹⁰.

¹⁰

https://github.com/nicolashernandez/nicolashernandez.github.io/blob/master/teaching/iutnantes_stage_FAQ.md

Responsabilités d'enseignement

Sur la période 3,5 - 4,5 responsabilités de modules avec préparation du matériel pédagogique, recrutement et coordination des intervenants, chargé des cours magistraux et d'au moins un groupe de TD

- Responsable du module DUT Info 1 "Architecture des Réseaux" (110 étu.)
- Resp. du module DUT Info 2 "Administration Système et Réseaux" (100 étu.)
- Co-resp. du module "Modélisation mathématique" en DUT Info 2 (80 étu.)
- Resp. du module de "Développement logiciel en industrie de la langue" en Master 2 Informatique - parcours ATAL (15 étudiants)

Jusqu'en 2019 Responsable du module "Outils de développement logiciel" en licence professionnelle MiAR (2*20 étudiants) ; avec intervention partielle en tant qu'enseignant

Les ressources pédagogiques complètes (consignes, vidéos, diaporamas, test d'auto-évaluation en ligne...) de deux de mes enseignements sont accessibles en ligne en accès ouvert :

- Architecture des réseaux¹¹ DUT 1ère année (M2102 du PPN¹² 2013)
- Administration des systèmes et des réseaux¹³ en DUT 2 (M4101C du PPN 2013)

Autres responsabilités pédagogiques

- *Participation annuelle au recrutement* des DUT 1ères années (3 jours pleins)
- *Participation annuelle aux portes ouvertes* du département INFO de l'IUT
- *Représentation du département INFO* à au moins un forum annuel d'orientation dans les formations du supérieur
- Depuis 2015 *participation aux International Days* organisés par l'enseignante d'anglais au département informatique en faisant cours en anglais
- 2020, accueil de 2 lycéens et présentation personnalisée de l'IUT et du département INFO dans le cadre du programme BRIO qui vise à **favoriser l'accès à l'enseignement supérieur à des jeunes** qui pour diverses raisons (contexte familial, territorial, milieux social et culturel modestes...) brident leur ambition professionnelle
<https://www.brio-nantes.org>
- 2019 *participation à un jury du Brevet de Technicien Supérieur* (BTS)
- 2016 *participation à des jurys de baccalauréats technologiques*

¹¹ <https://madoc.univ-nantes.fr/course/view.php?id=29848>

¹² <http://iut-informatique.fr/docs/ppn/fr.pdf>

¹³ <https://madoc.univ-nantes.fr/course/view.php?id=29877>