

Nicolas Hernandez

Maître de conférences

Nantes Université
LS2N (CNRS UMR-6004)

+33 (0) 2 51 12 53 94

✉ nicolas.hernandez@univ-nantes.fr

🌐 nicolashernandez.github.io



Rapport d'activités 2019–2022

Nom d'usage **HERNANDEZ** Prénom : **Nicolas**
Corps : **MCF** Grade : **MCF CN** Echelon : **8**
Discipline/section : **27 (Informatique)**

Ancienneté de 16 Ans, 3 Mois et 30 Jours dans le Corps et Grade et de 6 Mois et 15 Jours dans l'Echelon à la date du dépôt de la demande.

Synthèse du parcours professionnel et contexte d'exercice

Principales étapes de carrière

- 2021–aujourd'hui **Responsable de la future accréditation 2023-2028 du parcours ATAL**, *Master Informatique*, Dpt. Info., FST de Nantes Université
- 2021-2022 **Pilotage de la modification de maquette du parcours ATAL 2022/2023**, Dpt. INFO de l'IUT de Nantes, Dpt. Info., FST de Nantes Université
- 2021–aujourd'hui **Responsable du parcours ATAL (Apprentissage et Traitement Automatique des Langues)**, *Master2 Informatique*, Département Informatique, Faculté des Sciences et Techniques (FST) de Nantes Université
- 2013–2021 **Responsable des stages**, *DUT et Licence pro*, Dpt. INFO de l'IUT de Nantes
- 2010–2014 **Membre du Conseil Scientifique**, *IUT, Nantes*
- 2007–2010 **Membre élu au Conseil de Direction**, *IUT, Nantes*
- 2006–aujourd'hui **Maître de conférences CN**, *Nantes Université*
Rattaché au Dpt. INFO de l'IUT de Nantes et membre du Laboratoire d'Informatique Nantes Atlantique (LINA, UMR 6241) de 2006 à 2016 et du Laboratoire des Sciences du Numérique de Nantes (LS2N UMR CNRS 6004) depuis janvier 2017, équipe Traitement Automatique du Langage Naturel (TALN)
- 2005–2006 **Post-doctorat et vacataire**, *GREYC – Université de Caen*
- 2004–2005 **Demi-ATER**, *Institut d'Informatique d'Entreprise (CNAM-IIE)*, Evry
- 2003–2004 **Demi-ATER**, *Faculté des sciences et école d'ingénieur formation FIIFO de Paris-Sud*

2000–2003 **Moniteur**, *IUT d'Informatique d'Orsay*

Présentation de l'évolution de mes activités

Pour le département INFO de l'IUT de Nantes, mon recrutement en 2006 visait un renforcement de l'équipe pédagogique dans le pôle "architectures, systèmes et réseaux" (ASR). Mon implication dans les instances de mon établissement était réelle les 8 premières années ayant suivi mon recrutement (mandat au conseil scientifique et au conseil de direction). 2009 et 2012 correspondent aux années de naissance de mes enfants. De fait, mon investissement s'est réduit à cette période ou plutôt s'est développé autour d'une seule responsabilité : la gestion des stages et les relations avec les entreprises pour mon département.

Après 7 ans en tant que responsable des stages, en 2021, j'ai pris la coordination du parcours M2 ATAL (Apprentissage et Traitement Automatique des Langues) au sein du Master INFO de l'UFR Sciences. Ce parcours était alors chapeauté pour moitié par l'Université du Maine et offrait sa formation en initiale et en continue (avec alternance). Deux collègues à Nantes avaient en charge la coordination ; l'un partant à la direction du LS2N, il était entendu que nous poursuivions en binôme avec l'autre collègue. Les aléas de la vie ont fait que la collègue a bénéficié d'un congé parental sur l'année 2021-2022, et d'une CRCT en 2022-2023. Les collègues de mon équipe de recherche étant très pris par ailleurs, j'ai assumé la responsabilité du parcours seul. Il est prévu qu'à la rentrée 2023 nous coordonnions à nouveau le parcours en binôme ou que la collègue assume la charge seule pour me permettre de bénéficier à mon tour d'un congé de recherche. Les aléas ont aussi fait que la période fut intense. En raison d'un calendrier d'accréditation désynchronisée mais aussi de politiques de construction de pôles locaux forts en "Données et Intelligence Artificielle", la collaboration avec l'Univ. du Maine s'est terminée en 2021-2022. Outre la gestion "du quotidien" des initiaux et des alternants, il m'a fallu trouver une solution à base de mutualisation dans les masters locaux pour pouvoir ouvrir le parcours à la rentrée 2022, et cela, avec un calendrier compatible pour des alternants... Ces aspects ont été gérés sur fond de montage de la future accréditation 2023 du master INFO et de l'insertion du parcours ATAL dans une offre "Données et Intelligence Artificielle" à Nantes Université.

Au niveau de mon département de rattachement, la transformation du diplôme de DUT en BUT à la rentrée 2021 et le choix de mon département de n'ouvrir que le parcours "Réalisation d'applications : conception, développement, validation" dans un premier temps, font que des enseignements dans le pôle ASR où j'intervenais diminuent et se transforment. Le réinvestissement de mon énergie dans l'ouverture du parcours "Administration, gestion et exploitation des données" est à un stade de réflexion avec les collègues. Il constitue une opportunité car il se présente comme une passerelle possible vers les offres de Master auxquelles nous travaillons par ailleurs.

Formations

Académique

2004 **Doctorat en Informatique**, *Université de Paris-Sud XI*, Orsay, France, *Mention Très Honorable*

Sujet : Description et détection de structures de texte

Direction : Brigitte Grau et Christian Jacquemin (Lab. LIMSI/CNRS)

2000–2001 **DEA de Sciences Cognitives**, *Université de Paris-Sud XI, Assez-Bien*

1999–2000 **Maîtrise Informatique**, *Université d'Aix – Marseille II, Assez-Bien*
Erasmus - Université d'Uppsala (Suède)

Continue : expertise en apprentissage et traitement automatique des langues

Afin de me maintenir à niveau, je participe à des écoles sur mes thématiques de recherche.

2021 (45h) **Deep Learning School MOOC/SPOC**, *IVADO/Mila*

Machine Learning, Deep Learning and optimization ; Convolutional Neural Networks, Recurrent Neural Networks ; Bias and discrimination in AI

2016 (45h) **Machine Learning MOOC de Andrew Ng¹**, *Université de Stanford*, 95.7 %

Théorie et pratique, état de l'art en l'apprentissage supervisé, non supervisé ainsi que sur les bonnes pratiques à adopter

Continue : méthodologies d'encadrement

2021 (3j) **Mieux dans son job, mieux dans ses relations professionnelles**

Proposée par le LS2N autour de la Communication bienveillante

2019 (3j) **Diplôme d'Animateur Fédéral (DAF)**, *FFKDA*

Apprendre à organiser des entraînements sportifs de façon bénévole au sein des clubs affiliés à une fédération

2018, 2021 **Sauveteur et Secouriste du Travail SST**, *Nantes Université*
(3+1j)

2017 (28h) **Animer une équipe et piloter un projet**, *Pôle pédagogie" de l'UFR Sciences et Technique de l'Université de Nantes*

Apprentissage de techniques et jeux de rôles pour la gestion de réunion, gestion de situations conflictuelles, priorisation de tâches, construction et direction de projet (de formation pédagogique et scientifique), définition de critères de réussite et évaluation

Investissement pédagogique

1. Présentation synthétique de l'activité d'enseignement

En moyenne, ma charge annuelle de responsabilités pédagogiques et d'enseignement (encadrements compris) est de **369 heures ETD** (équivalent TD), avec une tendance à l'augmentation: 360H en 2018-2019, 380H en 2019-2020, 388H en 2020-2021, 397H en 2021-2022, 320H en 2022-2023 (prévisionnel).

Ce dépassement des heures statutaires n'est ni motivé par une nécessité sociale ni un désir personnel. De fait, l'ouverture de nos formations à l'alternance avec la création de groupes de TD supplémentaires, d'abord en 3e année, puis en 2e année, la nécessité en 2021-2022 de mener de front le DUT et le BUT, ont conduit à un volume accru d'heures au niveau du département. Jusqu'à présent, l'équipe pédagogique a réussi à absorber ce surplus mais la situation est de plus en plus tendue.

¹<https://coursera.org/share/a18472a3eb6cb1cfa19eaf35476df57f>

Le tableau suivant synthétise les temps consacrés aux activités déclarées sur une année :

Pôle	Responsabilité	SR	Compl.	Dev.	ATAL	Projet	Stage
	DUT/LP	L1/L2 CM/TD/TP	L2 CI/TD	L1/LP CI/TD	Master CM/TD/TP	L1/L2/LP/M	L2/LP/M1&2
ETD	46.50	104.29	28.33	47.43	31.85	43.70	49.5
%	13	30	8	13	9	12	14

- 4 thématiques : « Système et Réseau» (SR), « Développement logiciel» (Dev.), « Compléments à l'informatique e.g. mathématique» (Compl.), « Apprentissage et le Traitement Automatique des Langues (ATAL)»
- Près d'1/3 de mes heures ETD dans le pôle SR auprès des 1ère/2e année de BUT/DUT.
- ¼ concernent des activités de direction de projets ou de suivi d'étudiants (tutorat/stage).
- Mes pôles secondaires sont le Dev. et des enseignements Compl.
- J'ai la chance d'intervenir en MASTER dans mon domaine de recherche (ATAL).
- Pour cette période, mes responsabilités administratives furent de 2013 à 2021 la gestion des stages pour la 2e année de DUT et la Licence Professionnelle (LP) à l'IUT, et depuis 2021, la responsabilité d'un parcours du Master INFO à l'UFR Sciences ainsi que le montage de la nouvelle accréditation de celui-ci.
- 7 responsabilités : 5 Unités d'Enseignement (UE) et 2 interventions avec CM ou CI (cours intégrés déclarés en TD) avec la charge de constituer et diriger une équipe de 3 intervenants en DUT.
- En termes de création d'enseignements, la période COVID m'a conduit à réviser 2 UE de SR pour les rendre accessible à distance (vidéos, qcm en ligne...). En 2021-2022, la restructuration du master ATAL et mon invitation à participer au master CN m'ont conduit à recréer deux nouvelles UE en ATAL. La même année, l'ouverture du BUT 1 m'a conduit à réviser une UE de SR. Une année plus tard, en 2022-2023, la compréhension partagée du BUT se stabilise davantage ; je réviser à nouveau l'UE et l'étend. Globalement, mon rôle de responsable dans le M2 ATAL me conduit à co-diriger deux UE supplémentaires de 24H en termes de recherche et de programmation d'intervenants.

2. Présentation des enseignements

7 responsabilités d'UE/interventions en DUT/BUT, LP et Master.

Chargé des cours magistraux et d'au moins un groupe de TD, responsable de la préparation du matériel pédagogique, chargé du recrutement et de la coordination de 3 intervenants pour les UE en DUT/BUT.

Le détail incluant mes interventions en tant que chargé de TD et tuteurs peut être consulté en annexe.

Formation	Format	Matière	Effectif	H	Remarques
DUT Info 1	CM TD TP	M2102 Architecture des Réseaux	110/28/14	24h	Jusqu'en 2020-2021
BUT Info 1	CM TD TP	R2.04 Réseaux et communications bas niveaux	110/28/14	20h	2021-2022
BUT Info 1	CM TD TP	R2.05 Introduction aux services réseaux	110/28/14	40h	Depuis 2022-23
DUT Info 2	CM TD	M4101C Administration Système et Réseaux	88/22	24h	Jusqu'en 2021-2022
DUT Info 2	CM TD	M3202 Modélisation mathématique	88	24h	Jusqu'en 2021-2022
M2 INFO ATAL	CM TD TP	Dév. logiciel en industrie de la langue	12	24h	
M2 CN	CI	DATA 3	10	16h	depuis 2020-21 (4h en 2020-21)
LP MIAR	CM TD	Automatisation dev. ; Technologies XML	40	12h	Resp. d'interventions au sein d'UE

Analyse de mon activité d'enseignements

J'interviens dans la **formation de 3 diplômes et prioritairement (88%) sur des diplômes niveau Licence dans ma composante** (DUT/BUT et Licence pro). Une petite partie de mes heures sont en Master, ce qui pour moi est à la fois une opportunité et une nécessité pour l'enseignant chercheur que je suis.

Diplôme	DUT/BUT	Licence Pro.	Master
ETD/an	228.19	65.33	38.10
%	68.81%	19.70%	11.49%

Dans le détail des formations, **50% de mes interventions sont auprès de 2e année.** Cela vient pour bonne part du fait que mes enseignements sont majoritairement dans le pôle de compétences de l'ASR (Administration Système et Réseaux) dont plusieurs sujets sont avancés ou tout au moins requièrent des pré-requis avant d'être abordés. Notre département à l'IUT chapeaute une licence professionnelle (LP) dont la dénomination a évolué (SILR puis MIAR) mais qui globalement reste orientée développement logicielle. ATAL désigne un parcours du MASTER Info de la FST de l'université de Nantes. Il porte sur mon domaine de recherche : l'Apprentissage et Traitement Automatique des Langues. Les « Autres M2 » sont des interventions dans des masters ou des formations de niveau ingénieur.

Formation	DUT 1 INFO	DUT 2 INFO	LP MIAR/SILR	M1 ATAL	M2 ATAL	Autres M2
ETD/an	62.50	165.69	65.33	10.12	21.68	6.29
ETD %	18.85%	49.96%	19.70%	3.05%	6.54%	1.90%

Le tableau suivant présente la nature de mes interventions. La colonne « Responsabilité » correspond à ma charge de responsable des stages. La colonne « Encadrement » comprend « projets tutorés » et « suivi de stages ». **1/4 de mon temps est consacré à l'encadrement individuel ou de petits groupes. Plus de 42% de mes interventions sont des TD.** Mais bien que déclarées comme telles, **beaucoup correspondent à des Cours Intégrés (CI)** et résultent souvent de l'évolution des DUT à transformer des CM en TD et TP. Ma pratique pédagogique a évolué en conséquence, ainsi que mes supports de cours. Ceux-ci

mêlent désormais présentation théorique et manipulation pratique.

Nature	Responsabilité	CM	TD	TP	Encadrement
H. ETD/an	43.50	30.61	139.39	27.48	90.64
ETD %	13.12%	9.23%	42.03%	8.28%	27.33%

Le tableau suivant indique les **effectifs d'étudiants touchés selon les pôles de compétences**. Selon le pôle un même étudiant peut être compté plusieurs fois mais pour des cours différents. En DUT/BUT 1ère année, la promotion compte 110 étudiants, 88 en 2e année. La licence pro compte environ 40 étudiants. Le M1 et le M2 ATAL comptent respectivement 10 et 12 étudiants environ.

Pôle	Responsabilité	ASR	Compl. Dev.	TAL	Projet	Stage
Effectif/an	112.75	250.5	32.75	98.25	38.75	25.25 7.75

15-20% de mes heures sont dans des formations en alternance (contrat d'apprentissage en DUT 2e année et Licence professionnelle, contrats professionnels en Master 2).

Type de formation	initial	alternance	mixte
ETD/an	264.22	53.60	13.80
ETD %	79.67%	16.16%	4.16%

Un peu **plus de 88 % de mes interventions sont en présentiel**. **Près de 12% correspondent à du distanciel ou de hybride** (i.e. CM à distance et TD en présentiel ou bien CM/TD avec 1/2 des effectifs à distance). Cette **proportion est plus importante en Master (2/5 de mes ETD)**.

%ETD	distanciel	hybride	présentiel
Licence	5.34%	5.20%	89.45%
Master	4.27%	33.08%	62.64%
Total	4.17%	7.38%	88.45%

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 **177h (+ 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**).

3. Responsabilités pédagogiques

Mon engagement s'est traduit jusqu'à présent sous la forme d'une participation dans des instances administratives locales de l'IUT ainsi que dans les instances nationales universitaires et de l'éducation. Cette implication était réelle les 8 premières années ayant suivi mon

recrutement. 2009 et 2012 correspondent aux années de naissance de mes enfants. De fait, mon investissement s'est réduit depuis ou plutôt s'est développé autour d'une seule responsabilité : la gestion des stages et les relations avec les entreprises pour mon département de 2013 à 2021.

2021–... **Pilotage de la future accréditation 2023-2028 du parcours ATAL, Master Informatique**, Dpt. Info., FST de Nantes Université

Définition de la fiche métier, des compétences, des résultats d'apprentissage, ... en collaboration avec deux parcours colorés *IA* et *Données*, avec un objectif de mutualisation de ressources dans l'implémentation et d'ouverture en alternance. Constitution d'une équipe pédagogique et construction d'un programme M1-M2, recherche de responsables et d'intervenants. La tâche ne fut pas aisée la 1ère année car il s'agissait de travailler avec un établissement qui n'est pas mon établissement de rattachement avec des interlocuteurs et des modes de fonctionnement globalement différents.

2021-2022 **Pilotage de la modification de maquette du parcours ATAL 2022/2023**, Dpt. INFO de l'IUT de Nantes, Dpt. Info., FST de Nantes Université

Suite à une désynchronisation des calendriers d'accréditation entre NU et Le Mans Université avec laquelle les cours étaient mutualisés à 50% les années antérieures

2021–... **Co-responsable du parcours ATAL (Apprentissage et Traitement Automatique des Langues)**, *Master2 Informatique*, Département Informatique, Faculté des Sciences et Techniques (FST) de Nantes Université

Gestion de l'équipe pédagogique (notamment les extérieurs), de l'emploi du temps, recrutement, suivi des étudiants, stage, à la fois pour la formation initiale et l'alternance, supervision administrative de l'alternance avec le service dédié. En pratique charge exercée à 100% en raison d'un congé parental et d'un congé de recherche obtenu par ma collègue de 2021 à 2023

2013–2021 **Responsable des stages, DUT et Licence pro**, Dpt. INFO de l'IUT de Nantes

Relation avec les entreprises, formation et accompagnement des étudiants dans la recherche et l'exécution de leur stage, supervision du suivi des enseignants tuteurs, organisation des soutenances

Mais aussi : **Présidence de jurys** du Brevet de Technicien Supérieur (BTS) en 2019.

Participation à d'autres charges collectives telles que le recrutement annuel des DUT/BUT 1ères années (3 jours pleins de lecture et de classement de ± 150 dossiers sur $\approx 3\,000$ demandes), la journée annuelle de portes ouvertes du département, la journée annuelle de la recherche à l'IUT, la représentation du département INFO de l'IUT à au moins un forum annuel d'orientation dans les formations du supérieur, International Days organisés par l'enseignante d'anglais au département (cours du moment donnés en anglais).

4. Diffusion de la culture

Dans ce volet, mon action concerne essentiellement l'*ouverture* de ma discipline auprès de collègues, enseignants chercheurs, du domaine disciplinaire des sciences humaines et sociales (linguistes, psycho-linguistes, didacticiens).

En tant qu'informaticien d'abord puis comme enseignant et chercheur produisant des ressources pédagogiques, logicielles et langagières, je suis très attaché aux questions d'exploitation (utilisation, modification, diffusion, droit d'auteurs) des ressources produites. Ma vocation à travailler dans le public m'a conduit à m'intéresser notamment aux *licences dites "libres"*. Dans ce contexte, c'est naturellement qu'en tant que chercheur j'ai été amené à m'intéresser

aux enjeux et bonnes pratiques de la *science ouverte*.

En tant que rédacteur en chef de la revue Discours¹, j'ai ainsi beaucoup travaillé à la sensibilisation de mes collègues, membres du comité de rédaction, sur ces questions. J'ai notamment poussé à l'adoption et la mise en oeuvre des critères de bonnes pratiques éditoriales et de sciences ouvertes². Sur notre site web, il est désormais possible de lire notre engagement (politique de dépôt en archives ouvertes), et de constater notre transparence de fonctionnement via notre "Publication ethics and malpractice statement" ou les pages dédiées à l'explicitation de notre procédure éditoriale, du fonctionnement de nos comités et de notre procédure d'évaluation.

Une autre réalisation de ce brassage de cultures entre sciences numériques et sociales peut être observée au travers de l'obtention du soutien de l'ANR pour le projet PRC Lexhnology à l'interface des sciences du numérique et des sciences humaines et sociales (CE38) et qui réunit quatre laboratoires et trois disciplines à savoir l'informatique, la linguistique et la didactique.

5. Rayonnement, activités internationales

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

1. DUT Info 1 Architecture des réseaux M2102 2019-2022 (PPN 2013)
<https://madoc.univ-nantes.fr/course/view.php?id=29848>
2. DUT Info 2 Administration des systèmes et des réseaux M4101C 2019-2022 (PPN 2013)
<https://madoc.univ-nantes.fr/course/view.php?id=29877>
3. BUT Info 1 Réseaux et communications bas niveaux R2.04 2021-2022 (PN 2021)
<https://madoc.univ-nantes.fr/course/view.php?id=50785>
4. BUT Info 1 Introduction aux services réseaux R2.05 2022-2023 (PN 2021)
<https://madoc.univ-nantes.fr/course/view.php?id=50788>
5. Les « cahiers interactifs » (jupyter python notebook) combinant théorie et pratique de mes 2 enseignements en Master autour du Traitement Automatique des Langues 2021-2023 sont disponibles en ligne et en français
https://github.com/nicolashernandez/teaching_nlp

Le PPN représente le programme pédagogique national du DUT INFO et peut se consulter ici <http://iut-informatique.fr/docs/ppn/fr.pdf>. Le PN est le programme national du BUT INFO.

Autres ressources pédagogiques ouvertes et diffusées :

<https://nicolashernandez.github.io/#teaching>

1. 1 guide sur comment trouver un stage
2. 1 guide sur comment rédiger un rapport de stage ou de projet

Ressources diffusées en interne aux étudiants sur les stages (carte mentale, flux des tâches, FAQ, sondage, listes des entreprises ayant accueilli un stagiaire, consignes pour rédiger son rapport. . .) <https://uncloud.univ-nantes.fr/index.php/s/PLiN7C77sjk8Sy8> ; Mot de passe : 6tD7wojJ

¹discours.revues.org

²<https://www.inshs.cnrs.fr/fr/la-campagne-dattribution-de-subventions-aux-revues-scientifiques-de-linshs>

1. Présentation synthétique des thématiques de recherche

Ma discipline de recherche est fondamentalement le *Traitement Automatique des Langues* (TAL) avec une sensibilité sincère pour la *Linguistique* de par mon parcours et mes collaborations. Certains de mes travaux peuvent néanmoins être assimilés à de la *Fouille de Données Textuelles* (FdT). Mon domaine de recherche se situe dans le champ de la compréhension des textes par la machine et **mon objet d'étude est le "discours"**, comme unité textuelle cohérente supérieure à la phrase. Je m'intéresse en particulier aux **problématiques de modélisation et d'analyse/reconnaissance** de trois facettes du discours : 1) la caractérisation (d'aspects) d'*unités discursives*, 2) les *relations de cohérence* entre unités d'un même texte et 3) les *relations inter-textes*. Mes approches suivent l'évolution des paradigmes du TAL et portent aujourd'hui sur les questions de représentation de l'information discursive dans les modèles de langue neuronaux.

Mes travaux sur ces trois principaux thèmes se déclinent ainsi :

- **Aspects d'unités discursives** : lisibilité d'un texte [2], engagement des apprenants [1], représentation "profonde" des phrases [19, 20] ; projet ANR PASTEL et stage post-doctoral d'Amir Hazem et de M2 d'Esther Félix.
- **Structures discursives et de relations de cohérence** : identification du prochain énoncé, de la réponse à une question, désentremêlement de conversations, reconnaissance de parallélisme rhétorique, segmentation thématique [7, 17, 16] ; avec stage postdoctoral d'Amir Hazem.
- **Relations inter-textuelles** de similarités de contenu et de forme entre phrases [20, 19] ; projet ANR PASTEL et encadrement du stage post-doctoral d'Amir Hazem.

De manière secondaire ou transversale à mes collaborations, mes travaux portent aussi sur :

- La **production de ressources langagières** [2, 15] et **logicielles** [2] avec une attention particulière pour le traitement du français, la promotion de la science ouverte et l'apport d'une infrastructure commune de développement Avec Amir Hazem et Salima Mdhaffar ;
- L'**éducation comme domaine d'application** pour mesurer la complexité d'un écrit ou sa difficulté de lecture [2], pour mesurer l'engagement d'un apprenant [1], ou comme matière première [15, 22, 23]. ANR PRCE 2016 PASTEL, Labex CominLabs 2019 AILE, ANR PRC 2023 Lexhnology et thèse de Salima Mdhaffar ;
- Les **Communications écrites Médiées par les Réseaux** (CMR e.g. forum, chat, mails) comme matière première [1, 7, 17, 16] ; thèse de de Basma Boussaha.
- L'**évaluation des modèles et de la donnée** : mesure de l'adaptation au domaine d'un modèle [22, 23] ; avec Salima Mdhaffar.

2. Publications et productions scientifiques

Ma production scientifique compte environ 40 communications pour un peu plus de 300 citations et un h-index de 11 selon Google Scholar (<https://scholar.google.com/citations?user=SffWGZOAAAAJ>).

Sur la période, 11 communications dont 1 édition d'un numéro de revue et 1 conférence et 1

journal de rang A.

Felix, E., Hernandez, N., Rebaï, I., “Exploring the Limits of Lexicon-based Natural Language Processing Techniques for Measuring Engagement and Predicting MOOC’s Certification”. In: *14th International Conference on Computer Supported Education (CSEDU)*. Online Streaming, France, Apr. 2022

[Article long, session orale, conférence internationale (CORE B)]

Evaluation des performances sur des MOOC en Français d’une famille d’approches en TAL pour la mesure de l’engagement et la prédiction d’une certification

Boussaha, B. E. A. “End-to-End Response Selection Based on Multi-Level Context Response Matching”. In: *Computer Speech and Language* (2020)

[article long, revue internationale à comité de lecture (CORE A)]

Proposition d’une modélisation “profonde” hiérarchique du contexte dialogique (qui s’appuie à la fois sur la similarité des mots et de la séquence des mots) et de son intégration dans une approche “end-to-end” pour une tâche de prédiction du prochain énoncé

Mdhaffar, S. “Qualitative Evaluation of ASR Adaptation in a Lecture Context: Application to the PASTEL Corpus”. In: *Interspeech 2019*. Graz, Austria, Sept. 2019, pp. 569–573

[article long, session poster, conférences internationales (CORE A)]

Pour la tâche de reconnaissance de la parole, proposition d’une mesure d’évaluation des modèles de langues capable de mesurer le gain relatif à un domaine de connaissances.

Boussaha, B. E. A. “Towards Simple but Efficient Next Utterance Ranking”. In: *20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*. La Rochelle, France, Apr. 2019

[article long, session poster, conférence internationale (CORE B)]

Autour de la problématique de modélisation profonde des relations discursives entre deux énoncés, réflexion sur la possibilité de concevoir des architectures neuronales simples, moins gourmandes en ressources (énergie) mais toujours compétitives

Mdhaffar, S. “A Multimodal Educational Corpus of Oral Courses: Annotation, Analysis and Case Study”. In: *LREC*. Marseille, France, May 2020

[article long ; session orale conférence internationale (CORE C)]

Production d’un corpus pour soutenir l’étude de l’apport de la transcription automatique de la parole dans des contextes d’apprentissage humain en français. Ce corpus compile vidéos de cours magistraux, diaporama, alignement de temps de vidéo avec changement de page de diaporama, transcription manuelle, corpus du domaine extrait du web, annotation en mots clefs de la transcription/diaporama, annotation en concepts de la wikipedia, annotation en segments thématiques (2 grains) de la transcription

3. Encadrement doctoral et scientifique

- 3 encadrements de stage doctoral : 2 thèses soutenues (MESR et ANR), 1 en cours (MESR) et une 6^e à venir en 2023 (ANR)

- 2 encadrements de stage Master 2
- 4 encadrements stagiaires DUT, Licence et Master 1 (10 mois)
- 4 (groupes de) projets tutorés d'Étude et de Recherche en M1 INFO
- 4 (groupes de) projets tutorés de recherche en M2 INFO

4. Diffusion et rayonnement

- Membre de la société savante nationale ATALA et internationale ACL
- Depuis 2021, co-animateur du thème transverse "Technologie Numérique pour l'Éducation Ouverte" au sein du LS2N <https://www.ls2n.fr/theme/tneo>
- Depuis 2021, conseiller scientifique "Data Science et Intelligence Artificielle", auprès de CAPACITÉS (capacites.fr), filiale de valorisation de la recherche de Nantes Université

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, l'expertise de soumission de projets à l'ANR et de sujets de thèse CIFRE, la participation à des comités de suivi individuel de thèse.

- 13 comités de lectures de conférences
 - 2022, 2023 : Building Educational Application (BEA) workshop
 - 2013–2023 : Traitement Automatique des Langues Naturelles (TALN) ; Congrès franco-phonie 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
- 1 projet de Appel Générique de l'ANR : 2019 ; 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
- 1 thèse CIFRE 2019 ; expertise, pour l'Association nationale de la recherche et de la technologie (ANRT), de la qualité scientifique du projet, de la pertinence du contexte entrepreneurial, de l'adéquation du laboratoire, et du profil du candidat
- 3 Comités de Suivi Individuel (CSI) de thèse : Noé Cécillon, 2021–2023, Univ. d'Avignon, dir. Richard Dufour ; Elmokhtar Mohamed Moussa, 2020-2023, Nantes Univ., dir. Harold Mouchère ; Oriane Thiery 2022–2025, École Centrale de Nantes, dir. Diana Mateus

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

2019–aujourd'hui **Rédacteur en chef de la revue électronique multidisciplinaire (linguistique, psycholinguistique et informatique) Discours**, *réunion bimensuelle* recherche de financement (obtention de subventions bisannuelles INSHS de 1500€ en 2020 et 2022), 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, mise en place de l'*open sciences* <https://journals.openedition.org/discours>

Juin 2023 Co-éditeur du numéro varia 32 de la revue électronique Discours

Juin 2020 Co-éditeur du numéro varia 26 de la revue électronique Discours

Organisation colloques, conférences, journées d'étude

- 07 sept. Co-organisation avec Solen Quiniou d'une journée scientifique dans le GDR 2023 TAL sur thème « Les modèles de langage pour les domaines de spécialité »
- 2019 Organisation de deux 1/4 journées pour la restitution des travaux réalisés dans le cadre du projet structurant CominLabs AILE (AI for Learning Environments)

Ressources logicielles et langagières (développement, diffusion et maintenance)

Mes productions sont principalement diffusés via des dépôts github cf. <https://github.com/nicolashernandez?tab=repositories>. Ci-dessous une sélection.

- READI- (REAding DIfficulties) : développeur et directeur du dev.) module python 3 LREC22 pour évaluer la lisibilité d'un texte en français; licence Apache-v2. 1ère diffusion : juin 2022. <https://github.com/nicolashernandez/READI-LREC22>
- anr-pastel- data cours magistraux enrichis (transcription, diaporama, concepts...) : contributeur, et orchestrateur de la compilation et de la diffusion du corpus annoté produit durant le projet ANR PASTEL Project ; licence CC BY-NC-SA 4.0). 1ère diffusion : mars 2021. <https://github.com/nicolashernandez/anr-pastel-data>

5. Responsabilités scientifiques

Projets financés avec participation au montage et pilotage

- 2023 (42 mois) **Lexhnology (*Modélisation linguistique et computationnelle de la structure discursive des textes juridiques appliquée à l'apprentissage des langues*)**
 - Soutien ANR PRC CE38 (Interfaces : sciences du numérique – sciences humaines et sociales) - 02/01/2023 42 mois Subvention : 245 040,50 € sur 533 426,68 €
 - Partenariat CRINI (Porteur), LS2N, ATILF, LAIRDIL
 - Mission représentation de l'information discursive des rôles rhétoriques dans les modèles de langage neuronaux; développement d'un analyseur de la structure argumentative d'une langue de spécialité (jurisprudences américaines) ; aide à l'apprentissage d'une langue de spécialité pour des apprenants L2
 - url <https://lexhnology.hypotheses.org/>
- 2019 (10 mois) **AILE (*Artificial Intelligence for Learning Environments*)**
 - Soutien Labex CominLabs "projet exploratoire" – Montant alloué 5 675 €
 - Partenariat Co-porteur avec Olivier Aubert ; LS2N (équipes Duke et TALN), IMT Atlantique, Centre de développement pédagogique de l'Univ. de Nantes

Mission Rassembler des chercheurs et consolider l'écosystème régional autour des questions de l'intelligence artificielle et de l'éducation, entre autres : organisation de 2 ½ journées de séminaires thématiques, financement de stages, organisation d'un atelier national, exposés des travaux réalisés, foyer de montage d'au moins deux projets de recherche multipartenaires

url <http://aile.comin-ocw.org>

2016–2020 **PASTEL (*Transcription Automatique de la Parole pour l'Apprentissage et la Formation*)**
(54 mois)

Soutien ANR/PRCE – Subvention : 98 118 € sur 686 242 €

Partenariat LIUM (porteur), CREN, Orange Lab, LS2N

Mission 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 contrôlée et alignement de ressources pédagogiques

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

En 2020, en partenariat avec IMT Atlantique/Lab-STICC, INALCO et le Service Universitaire des Langues "Français Langue Etrangère" de Nantes Université, soumission du projet MISO (Measuring Interaction Supporting Online Education) à la filière "IA pour la Formation" du Labex CominLabs MISO (Measuring Interaction Supporting Online Education), accepté pour soumission en version étendue mais non retenu au second tour. Le projet avait pour objet de mesurer l'engagement des apprenants afin de soutenir l'apprentissage dans un contexte d'enseignement en ligne massif. En 2021, soumission à l'ANR d'une version modifiée qui n'a pas été retenue non plus.

Prix, bourses ou levées de fond externes

2020, 2022 **Obtention de deux subventions bisannuelles INSHS**, Montant 1500 €,

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

Financement relecture de l'anglais, réunion du comité rédactionnel

Pour l'équipe de recherche TALN@LS2N

Responsable de différentes actions structurantes autour de l'infrastructure logicielle et matérielle de l'équipe (notamment avant l'avènement de l'ère de github et d'huggingface)

Depuis 2016 Soutien à l'activité deep learning : analyse des besoins de l'équipe, identification du matériel requis, devis, étude et expérimentation d'alternatives dans l'écosystème nantais, achat et supervision de l'exploitation de machines GPU, maintenance

1. Présentation synthétique des responsabilités

Membre de conseils administratifs et scientifiques

- De 2010 à 2014, Conseil Scientifique, IUT, Nantes
- De 2007 à 2010, membre élu au Conseil de Direction, IUT, Nantes

Membres de jurys de formations, des comités de recrutement, de poursuite d'étude, des commissions mixtes étudiants/enseignants

- De 2006 à 2022, DUT Informatique, et depuis 2021, BUT Informatique, IUT, Nantes
- Depuis 2019, Master INFO de l'UFR S&T, parcours ALMA/ORO/ATAL, Nantes

2. Responsabilités administratives

Mon tempérament m'amène à préférer les responsabilités opérationnelles. Le métier rentrant avec les ans, je ne m'interdirais pas de m'investir dans la direction d'une structure de mes établissements d'enseignement ou de recherche.

Néanmoins pour les 3 prochaines années, mon projet est davantage de m'investir dans le montage de la formation du BUT et de l'implémentation de son parcours "administration, gestion et exploitation des données".

3. Responsabilités et mandats locaux ou régionaux

Participation à la charge collective du département telles que le recrutement annuel des DUT/BUT 1ères années (3 jours pleins d'analyse de dossiers), la journée de porte ouverte du département, la journée recherche de l'IUT, la représentation du département INFO à au moins un forum annuel d'orientation dans les formations du supérieur, International Days organisés par l'enseignante d'anglais au département (cours donnés en anglais).

4. Responsabilités et mandats nationaux ou internationaux

Une expertise de projet PRC pour l'ANR en 2019, une expertise d'un dossier CIFRE pour l'ANRT en 2019, une présidence de jurys du Brevet de Technicien Supérieur (BTS) en 2019.

1. Tableau des enseignements

Les diplômes LP SIL et MIAR font référence à différentes habilitations de la licence professionnelle (LP) orientée développement logiciel, chapeauté par le dpt INFO de l'IUT. Le Master CN signifie Culture Numérique et est chapeauté par l'école d'ingénieur Polytech.

(1) Type de formation (i)nitiale / initiale en (a)lternance / (c)ontinue professionnelle, (p)résentielle / à (d)istance ; (2) CM, TP, TD, encadrement de travaux de fin d'étude, de stages et de projets tutorés ; (3) Volume horaire annuel Equivalent TD. Le total annuel approximatif ne compte pas les heures de responsabilités pédagogiques.

Année	Niveau	Diplôme	Intitulé	Type (1)	Nature (2)	Effectifs	ETD (3)
2018/2019 (≈328h)	L1	DUT INFO	Introduction à l'algorithmique	i/p	TP	28	16
			Tutorat	i/p	Encadrement	7	2.67
	L2	DUT INFO	Architecture des réseaux	i/p	CM TD TP	110/28/14	38
			Services réseaux	i/p	TD		22
			Modélisations mathématiques	i,a/p	TD	44	46.67
			Projet tutoré	i/p	Encadrement	20	30
			Administration système et réseau	i,a/p	CM TD	88/22	60
			Stage	i/p	Encadrement	5	15
	L3	LP MIAR	Collaborer en anglais	a/p	Encadrement	3	1.33
			Technologies web côté client	i/p	TD	22	16
			Outils pour le développement	i/p	TD	22	16
			Stage	i,a/p	Encadrement	3	43
	M1	Master INFO ATAL	Initiation à la recherche	i/p	Encadrement	3	3
	M2	Master INFO ATAL	Interaction and applications	i/p	CM	12	2
Stage			i/p	Encadrement	1	3	
			Dév. logiciel en industrie de la langue	i/p,d	CM TD TP	24	18
2019/2020 (≈330h)	L1	DUT INFO	Introduction à l'algorithmique	i/p	TD TP	28	20
			Tutorat	i/p	TP	7	1.33
	L2	DUT INFO	Architecture des réseaux	i/p	CM TD TP	110/28/14	49.65
			Services réseaux	i/d	TD		22
			Modélisations mathématiques	i,a/p	TD	22	32
			Projet tutoré	i/p	Encadrement	20	30
			Administration système et réseau	i,a/d	CM TD TP	88/22	59
			Stage	i/p	Encadrement	2	6
	L3	LP MIAR	Technologies web côté client	i,a/p	TD	44	16
			Outils pour le développement	i,a/p	TD	44	16
			Stage	i,a/p	Encadrement	3	43
			Initiation à la recherche	i/p	Encadrement	3	3
	M1	Master INFO ATAL	Interaction and applications	i/p	CM	12	2
			Stage	i,c/p	Encadrement	3	9
M2	Master INFO ATAL	Conférences et cours invités	i/p	CM	12	2	
		Dév. logiciel en industrie de la langue	i/p,d	CM TD TP	24	20.88	

Année	Niveau	Diplôme	Intitulé	Type (1)	Nature (2)	Effectifs	ETD (3)
2020/2021 (≈301h)	L1	DUT INFO	Tutorat	i/p	TD	8	13.33
			Architecture des réseaux	i/p	CM TD TP	110/28/14	50
	L2	DUT INFO	Services réseaux	i/p	TD	22	22.5
			Modélisations mathématiques	i,a/p	TD	44	32
			Collaborer en anglais	i/p	Encadrement	22	1.33
			Projet tutoré	i/p	Encadrement	20	20
			Administration système et réseau	i,a/p	CM TD TP	88/22	36.18
			Stage	i,a/p	Encadrement	6	17
	L3	LP MIAR	Technologies web côté client	a/p	TD	44	16
			Outils pour le développement	a/p	TD	44	16
			Stage	a/p	Encadrement	3	43
	M1	Master INFO ATAL	Initiation à la recherche	i/p	Encadrement	12	1
	M2	Master INFO ATAL	Interaction and applications	i/p	CM	12	2
Stage			i/p	Encadrement	2	6	
M2	Master CN	Dév. logiciel en industrie de la langue	i/p	CM TD TP	24	12.44	
		Projet tutoré	i/p	Encadrement	6	4	
			Introduction au TAL	i/p	CM TD	10	4.5
2021/2022 (≈397h)	L1	BUT INFO	Tutorat	i/p	Encadrement	7	1.33
			SAÉ Installation poste	i/p	TD	28	6.67
			Réseau et bas niveau	i/p	CM TD TP	110/28/14	26
			Services réseaux	i/p	TD TP	28/14	18.66
	L2	DUT INFO	SAÉ Exploration d'algorithmes	i/p	TD	28	11
			Services réseaux	i/p	TD	22	20
			Modélisations mathématiques	i/p	TD	22	29.33
			Projet tutoré	i/p	Encadrement	4	12
			Administration système et réseau	i,a/p	CM TD TP	88/22	51.33
			Projet tutoré de développement logiciel	i/p	Encadrement	4	8
	L3	LP MIAR	Stage	i,a/p	Encadrement	6	35
			Technologies web côté client	i/p	TD	44	16
			Outils pour le développement	i/p	TD	44	16
	M1	Master INFO ATAL	Stage	a/p	Encadrement	3	60
			Initiation à la recherche	i/p	Encadrement	3	2
			Interaction and applications	i/p	CM	12	2
	M2	Master INFO ATAL	Dév. logiciel en industrie de la langue	i/p,d	CM TD TP	12	22.45
			Projet tutoré	i/p	Encadrement	2	3
M2	Master CN	Stage	i,a/p	Encadrement	3	36	
		Introduction au TAL	i/p	CM TD	10	20.75	
2022/2023 (≈320h)	L1	BUT INFO	Tutorat	i/p	Encadrement	7	1.33
			Services réseaux	i/p	CM TD TP	110/28/14	40
	L2	BUT INFO	SAÉ Configuration des Services Réseaux	i/p	TD	28	15
			Architecture des réseaux	i/p	TD	22	10.64
			Cryptographie et Sécurité	i/p	TD	22	10.64
			Développement d'une application complexe	i/p	TD	22	8
			Méthodes d'optimisation	i/p	TD	22	10.64
			Automate et langage	i/p	TD	22	10.64
	L3	LP MIAR	Stage	i,a/p	Encadrement	6	35
			Technologies web côté client	i/p	TD	22	16
			Outils pour le développement	i/p	TD	22	16
	M1	Master INFO ATAL	Stage	a/p	Encadrement	3	60
			Initiation à la recherche	i/p	Encadrement	3	2
			Interaction and applications	i/p	CM	12	2
	M2	Master INFO ATAL	Dév. logiciel en industrie de la langue	i/p,d	CM TD TP	12	22.45
Projet tutoré			i/p	Encadrement	2	3	
M2	Master CN	Stage	i,a/p	Encadrement	3	36	
		Introduction au TAL	i/p	CM TD	10	20.75	

2. Liste classée des publications

Sur la période, 11 communications dont 1 édition d'un numéro de revue et 1 conférence et 1 journal de rang A.

Ma production scientifique totale compte environ 40 communications pour un peu plus de 300 citations et un h-index de 11 selon Google Scholar (<https://scholar.google.com/citations?user=SffwGZ0AAAAAJ>).

Le rang des conférences ci-dessous suit la classification Computing Research and Education Association of Australasia³ (CORE).

Activité éditoriale

- [10] Nicolas Hernandez and Lydia Mai Ho-Dac, eds. *Discours, revue électronique de linguistique, psycholinguistique et informatique*. Vol. 26. Varia. Presse Universitaire de Caen, June 2020.

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

- [7] Boussaha, B. E. A., Hernandez, N., Jacquin, C., Morin, E., "End-to-End Response Selection Based on Multi-Level Context Response Matching". In: *Computer Speech and Language* (2020).

Conférences internationales (CORE A)

- [22] Mdhaffar, S., Estève, Y., Hernandez, N., Laurent, A., Dufour, R., Quiniou, S., "Qualitative Evaluation of ASR Adaptation in a Lecture Context: Application to the PASTEL Corpus". In: *Interspeech 2019*. Graz, Austria, Sept. 2019, pp. 569–573.

Conférences internationales (CORE B)

- [1] Felix, E., Hernandez, N., Rebaï, I., "Exploring the Limits of Lexicon-based Natural Language Processing Techniques for Measuring Engagement and Predicting MOOC's Certification". In: *14th International Conference on Computer Supported Education (CSEDE)*. Online Streaming, France, Apr. 2022.
- [17] Boussaha, B. E. A., Hernandez, N., Jacquin, C., Morin, E., "Towards Simple but Efficient Next Utterance Ranking". In: *20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*. La Rochelle, France, Apr. 2019.

Conférences internationales (CORE C)

- [15] Mdhaffar, S., Estève, Y., Laurent, A., Hernandez, N., Dufour, R., Charlet, D., Damnati, G., Quiniou, S., Camelin, N., "A Multimodal Educational Corpus of Oral Courses: Annotation, Analysis and Case Study". In: *LREC*. Marseille, France, May 2020.
- [19] Hazem, A., Hernandez, N., "Meta-Embedding Sentence Representation for Textual Similarity". In: *Recent Advances In Natural Language Processing (RANLP)*. Varna, Bulgaria, Sept. 2019.
- [20] Hazem, A., Hernandez, N., "Tweaks and Tricks for Word Embedding Disruptions". In: *Recent Advances In Natural Language Processing (RANLP)*. Varna, Bulgaria, Sept. 2019.

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

³<http://www.core.edu.au/conference-portal>

- [23] Mdhaffar, S., Estève, Y., Hernandez, N., Laurent, A., Quiniou, S., “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: *Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN) - PFIA 2019 - Volume II : Articles courts*. Ed. by Emmanuel Morin, Sophie Rosset, and Pierre Zweigenbaum. Toulouse, France: ATALA, July 2019, pp. 167–174.

Ateliers internationaux (avec comité de sélection et actes)

- [2] Hernandez, N., Faine, T., Oulbaz, N., “Open corpora and toolkit for assessing text readability in French”. In: *Proceedings of the 2nd Workshop on Tools and Resources to Empower People with REAding Difficulties (READI) within the 13th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, June 2022, pp. 54–61.
- [16] Boussaha, B. E. A., Hernandez, N., Jacquin, C., Morin, E., “Multi-level Context Response Matching in Retrieval-Based Dialog Systems”. In: *Dialog System Technology Challenges Workshop at AAAI*. Honolulu, United States, Jan. 2019.

3. Liste des directions et codirections de thèses, post-doctorats...

3 encadrements de stage doctoral

Au total 5 encadrements et sur la période 2 thèses soutenues et 1 en cours.

2022–... **Léane Jourdan**, Financement MESR

Sujet Neural approaches for modelling and analysing the argumentative structure of Research articles

Encadrement R. Dufour 40%, **N. Hernandez 30%**, F. Boudin 30%

2017-2020 **Salima Mdhafar**, 40 mois, Financement ANR

Sujet Enrichissement de modèles avec des documents extérieurs

Encadrement Y. Estève[†] 25%, **N. Hernandez*** 25%, A. Laurent[†] 25%, S. Quiniou* 25% – [†]LIUM, *LS2N

Publications 1 CORE A ; 1 CORE C

Devenir Post-doctorat

2016-2019 **Basma El Amel Boussaha**, 38 mois, Financement MESR

Sujet Response Selection for End-to-End Retrieval-Based Dialogue Systems

Encadrement E. Morin 40%, **N. Hernandez 30%**, C. Jacquin 30%

Publications 1 CORE B ; 3 CORE C

Devenir Senior data scientist/NLP chez della.ia (startup)

Encadrements de stage post-doctoral

27 mois d'encadrement post-doctoral réparties sur 3 personnes différentes de 2008 à 2017 mais aucun sur la période.

3 encadrements de stage de Master 2

Anas Belfathi – Février 2023 (6 mois) – Financement ANR Lexhnology

Sujet *Diving into neural language models for improving discourse analysis tasks*

Encadrement N. Hernandez 50 %, Laura Monceaux 50 % (LS2N)

Esther Félix – Février 2021 (6 mois) – Financement LS2N

Sujet Apport du TAL pour l'assistance à l'apprentissage humain - mesures de l'engagement des apprenants dans les forums de MOOCs

Encadrement N. Hernandez 80 %, Issam Rebaï 20 % (IMT Atlantique)

Publications 1 CORE B

Samy Gascoin-Fontaine – Octobre 2020 (3 mois) – Financement ANR PASTEL

Sujet Valorisation - paquetage de données collectées et produites durant le projet ANR PASTEL

4 encadrements stagiaires DUT, Licence et Master 1

Nabil Oulbaz – Avril 2022 (2,5 mois) – Master Info 1 – Co-financé : LS2N+IUT de Nantes

Sujet Impact du genre des données en pretraining, fine-tuning et test dans le cadre d'une tâche de classification d'émotions

Tristan Faine – Avril 2022 (2,5 mois) – Master Info 1 – Financement : Equipe TALN

Sujet Implémentation d'une bibliothèque logicielle libre pour évaluer la lisibilité d'écrits en français

url <https://github.com/nicolashernandez/READI-LREC22>

Axel Rambaud – Avril 2019 (2,5 mois) – Master Info 1 – Financement : ANR PASTEL

Sujet 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

Sujet Définition et expérimentation de mesure automatique de la notion de difficulté et de concepts difficiles dans un cours

Autres encadrements (projets M1 et M2 informatique)

4 (groupes de) *Travaux d'Étude et de Recherche en M1 INFO*

≈ 1h hebdo de suivi sur un sujet proposé visant une initiation à la recherche. Les étudiants y consacrent 1 jour/semaine sur 14 semaines (3,5 mois)

- 2021–2022 **Nabil Oulbaz et Tristan Faine**
Evaluer la lisibilité d'un texte ou sa complexité linguistique
- 2020–2021 **Quentin Daëron et Sylouan Corfa**
Learner Engagement analysis in MOOCs' forums
- 2019–2020 **Aniss Bentebib, Amine Boulahmel, Harry Jandu**
Détection et analyse de sentiments à base d'aspects dans un corpus français
- 2018–2019 **Maxime Loizon, Axel Rambaud, Zina Rezgui**
Annotating Course Materials by Wikipedia Concepts

4 (groupes de) *Projet de recherche en M2 INFO*

A raison d'1 à 2h hebdomadaire en moyenne de suivi sur un sujet de recherche proposé. Les étudiants y consacrent 120h sur 4-7 semaines à temps plein.

- 2021–2022 **Othmane Chentouf, Ronan Belleil et Séraphin Costa**, (co-encadrement avec *Richard Dufour*)
Impact de position des mots dans la construction des modèles de langues figés et contextuels
- 2020–2021 **Oumaima El-Khettari, Fatima Er-Rafay, Naima Mazari, Marc Le Bras**
Anonymisation des conversations écrites en ligne
- 2019–2020 **Thomas Fonteneau, Maxime Loizon, Samy Gascoin-Fontaine**
Apprentissage supervisé avec des données artificielles
- 2019–2020 **Mohamed Reda Marzouk, Axel Rambaud**
De l'abstraction des mots à la difficulté des ressources