

# Curriculum Vitae

*Docteur en NLP (Natural Language Processing), Sorbonne Université*

**Dhaou GHOUL**

159 Rue du Général Leclerc 94000 Créteil

Tel : 06 52 75 98 33

E-mail : [Dhaou.ghoul@gmail.com](mailto:Dhaou.ghoul@gmail.com)

Page : <http://cereli.fr/membres/dhaou-ghoul/>

*Double nationalité : française  
et tunisienne*

*Né le 27/12/1984 à Medenine,  
Tunisie*

## Diplômes

---

<b>Qualification aux fonctions de maître de conférences</b> Section Informatique (27) du CNU	2019
<b>Thèse de doctorat</b> en sciences du langage, spécialité informatique <i>Classifications et grammaires des invariants lexicaux arabes en prévision d'un traitement informatique de cette langue.</i> Sorbonne Université.	2016
<b>Master recherche, sciences du langage / industries de la langue (IDL)</b> , Spécialité: Traitement automatique de la langue Écrite Université Grenoble Alpes, France	2011
<b>Maîtrise en Informatique</b> Université de Monastir, Tunisie	2008
<b>Diplôme de premier cycle Spécialité Mathématique-Informatique</b> Université de Monastir, Tunisie	2006

## Expériences professionnelles

---

<b>ATER, UR 4509, STIH Sorbonne Université, Paris</b> Linguistique de corpus et analyse de données textuelles: Recherche et Extraction d'informations, fouille de textes, Classification de documents, Classification des tweets.	Septembre 2018- Août 2020 Sorbonne Université
<b>Institut français d'archéologie orientale (IFAO), Le Caire, Égypte</b> Test d'un outil de traitement automatique de la langue arabe, « <i>Kawâkib</i> » dans le cadre du projet Mogador (Modélisation des grammaires arabes, des données et des outils de recherche).	Février 2016- Avril 2016 IFAO, Égypte
<b>UR 4509, STIH, Sorbonne Université, Paris</b> <i>Recherche en Linguistique-Informatique:</i> Classifications et grammaires des invariants lexicaux arabes en prévision d'un traitement informatique de cette langue : cas des invariants lexicaux temporels	Décembre 2012- Novembre 2016 Sorbonne Université
<b>Entreprise ETL express</b> Coordinateur Web	Janvier 2017/ Mars 2017 Grenoble, France.

**UR 609, LIDILEM, Université de Grenoble Alpes, France**

*Mission TAL : Elaboration d'outils génériques pour l'étiquetage morphosyntaxique de la langue arabe : segmentation et définition d'un corpus d'entraînement* Février 2011- Septembre 2011  
LIDILEM, Grenoble

**Laboratoire RIADI, Université de Monastir, Tunisie** Mars 2008- Juin 2008  
*TAL : Réalisation d'un étiqueteur grammatical pour la langue arabe* Monastir, Tunisie

## **Expériences dans l'enseignement**

---

**Septembre 2018 – Août 2020** : ATER en informatique, Sorbonne Université Paris, France.

**Avril 2017 - Juillet 2018** : Enseignant en Mathématiques, Collège Pré-Bénit, Bourgoin Jallieu, Isère.

**Mars 2017** : Enseignant en Mathématiques au Lycée Les eaux Claires, Grenoble.

**2015/2016** : Enseignant vacataire en Théories des automates à l'université de Monastir.

**2015/2016** : Enseignant vacataire en C2i à l'université de Monastir.

**2014/2015** : Enseignant vacataire en bases de données à l'université de Monastir.

**2013/2014** : Enseignant vacataire en atelier de Programmation à l'université de Monastir.

## **Compétences informatiques**

---

<b>Langages de programmation</b>	Python, Java, Perl, R
<b>Développement Web</b>	HTML, PHP, CSS, JavaScript
<b>Bureautique</b>	Microsoft office, open office, Latex, Excel.
<b>Modélisation</b>	Génie logiciel, Merise, UML.
<b>Base de données</b>	Access, MySQL, SQLWorkbench
<b>Outils TAL/ Textométrie</b>	spaCy, nltk, TRAMEUR, TXM
<b>Divers</b>	Linux, Git, scikit-learn; Tensorflow, Docker, Spark

## **Activités scientifiques**

---

**Janvier 2019 - Juin 2019** Co-encadrement d'un mémoire (master)

**Sujet** : Analyse et classification d'opinions dans des tweets arabes.

**Mars 2019 - Août 2019** Participation au DEFT (Défi fouille de textes) et participation à la compétition MADAR (Identification du dialecte arabe), WANLP 2019: The Fourth Workshop for Arabic Natural Language Processing

**2016/2017** Co-éditeur du volume UNIGRA avec Claire Martinot, Paris, CRL.

**2013-aujourd'hui** Membre actif à la Cellule de recherche en linguistique (CRL).

**Décembre 2016** Membre du comité d'organisation du colloque international *Les traductions françaises du coran et de ses exégèses*, Université Paris-

Sorbonne, France.

- Juin 2014** Membre du comité d'organisation du colloque international *La langue, la voix, la parole*, Sorbonne Université, France.
- Août 2015** Membre du comité d'organisation du colloque international *Universalité et grammaire: paradoxe insoluble ou solution matricielle ?* (UNIGRA), UBFC, Besançon, France.
- 2013-2015** Participation au séminaire « Invariances et calculabilité en langue arabe et en sémitique » (ICLAS), MMSH, Aix-en Provence, France.

## Langues

---

Arabe : langue maternelle, Français : Bilingue, Anglais : Avancé.

## Références

---

- Gaël Lejeune, Maître de Conférences en Informatique, Sorbonne Université, [gael.lejeune@sorbonne-universite.fr](mailto:gael.lejeune@sorbonne-universite.fr).
- Olivier Kraif, Maître de Conférences en Informatique, Université Grenoble Alpes, [Olivier.Kraif@univ-grenoble-alpes.fr](mailto:Olivier.Kraif@univ-grenoble-alpes.fr).
- Claire Martinot, Professeur des universités en sciences du langage, Sorbonne Université, [claire.martinot@espe-paris.fr](mailto:claire.martinot@espe-paris.fr).

## Publications

Depuis le début de ma carrière de chercheur, j'ai publié quinze articles à partir de conférences internationales et un article à partir d'une conférence nationale.

## Ouvrages scientifiques (ou chapitres d'ouvrages) (OS) :

- **Dhaou Ghoul**, 2017. «Identification et attentes syntaxiques des invariants lexicaux temporels de la langue arabe ». *Universalité et grammaire : paradoxe insoluble ou solution matricielle?* Claire MARTINOT & Dhaou GHOUL (éds). Paris, CRL, Mars 2017, pp 182-192, ISBN : 9782952602785, <http://www.sudoc.abes.fr/xslt/DB=2.1//SRCH?IKT=12&TRM=199386080>.

## Conférences CORE A\* :

- **Dhaou Ghoul**, Gaël Lejeune « MICHAEL: Mining Character-level Patterns for Arabic Dialect Identification (MADAR Challenge). » Proceedings of the Fourth Arabic Natural Language Processing Workshop, p. 229-233 (Workshop de ACL 2019).

## Conférences CORE A :

- **Dhaou Ghoul**, Gaël Lejeune « Sarcasm and Sentiment Detection in Arabic:investigating the interest of character-level features. » Proceedings of the Sixth Arabic Natural Language Processing Workshop (WANLP 2021 @EACL2021), à paraître.

- **Dhaou Ghoul**, Gaël Lejeune « Comparison between Voting Classifier and Deep Learning methods for Arabic Dialect Identification. » Proceedings of the Fifth Arabic Natural Language Processing Workshop (WANLP 2020 @COLING2020), (Workshop de COLING, 9ème place /18).

### **Conférences CORE B :**

- **Dhaou Ghoul**, Gaël Lejeune, Lichao Zhu « Evaluating Lightweight text classification and Information Extraction for Arabic texts ». CicLing'19. The 20th international conference on intelligent text processing and computational linguistics, La Rochelle, France, 07-13 Avril 2019
- **Dhaou Ghoul** « Modelling through a grammar diagram of the complex Arabic lexical invariants: "the temporal invariants" ». ICEMIS'17, International Conférence on Engineering and MIS 2017, Monastir, Tunisia, 8-10 May 2017. ([http://www.iares.net/images/conference%20program\\_updated.pdf](http://www.iares.net/images/conference%20program_updated.pdf))
- **Dhaou Ghoul**, André Jaccarini, Amr Helmy Ibrahim "Classifications and grammars of simple Arabic lexical invariants in anticipation of an automatic processing of this language: "the temporal invariants". ". CicLing'17, The 18th international conference on intelligent text processing and computational linguistics, Budapest, Hungary, 17-23 Avril 2017. (<https://www.cicling.org/2017/accepted.html>)

### **Conférences CORE C :**

- Davide Buscaldi, Ghazi Felhi, **Dhaou Ghoul**, Joseph Le Roux, Gaël Lejeune and Xudong Zhang « Calcul de similarité entre phrases : quelles mesures et quels descripteurs ? » DEFT@JEP/TALN/RECITAL 2020, p. 14-25 (2ème place sur deux tâches du défi).
- Davide Buscaldi, **Dhaou Ghoul**, Joseph Le Roux and Gaël Lejeune « Indexation et appariements de documents cliniques pour le Deft 2019. » Atelier DEFT 2019, p. 49-56 (Workshop de TALN).
- **Dhaou Ghoul**, Amr Helmy Ibrahim, Claude Audebert "Rules-based grammatical and semantic disambiguation of the token “ḥattā” in Arabic". ICTA'2015, 5th international conference on information and communication Technology and Accessibility, Marrakech, Maroc, 21-23 December 2015 (<https://ieeexplore.ieee.org/document/7426889>).
- **Dhaou Ghoul**, Amr Helmy Ibrahim, "Web Arabic corpus: Construction d'un large corpus arabe annoté morpho-syntaxiquement à partir du Web". CEC-TAL'2015, Colloque pour les Étudiants Chercheurs en Traitement Automatique du Langage naturel et ses applications, Sousse, Tunisie, 23-25 Mars 2015, pp 12-16. ([https://web2.qatar.cmu.edu/~wajdiz/cectal2015/Actes\\_CEC-TAL2015.pdf](https://web2.qatar.cmu.edu/~wajdiz/cectal2015/Actes_CEC-TAL2015.pdf)).
- **Dhaou Ghoul**, "Construction d'un corpus arabe à partir du Web dans le but d'identifier les mots-outils ou tokens". JADT'14, Journées International d'Analyse Statistiques des données textuelles, Paris, France, 2-6 juin 2014, pp 271-276 (<http://lexicométrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/22-JADT2014.pdf>).

- **Dhaou Ghoul**, "Développement de ressources pour l'entraînement et l'utilisation de l'étiqueteur morphosyntaxique TreeTagger sur l'arabe". Conférence TALN-RECITAL, Sables d'Olonne, France, 17-21 juin 2013 pp 69-82 (<http://aclweb.org/anthology/F/F13/F13-5006.pdf>).
- Mohamed Ahref Ben Mohamed, **Dhaou Ghoul**, Mohamed Amine Nahdi, Mourad Mars, Mounir Zrigui "Arabic CALL system based on pedagogically indexed text". IC-AI'11 - The 2011 International Conférence on Artificial Intelligence. Las Vegas, Nevada, USA, 18-21 juillet 2011 (<https://arxiv.org/abs/1207.2373>).

### **Communications orales sans actes dans un congrès national:**

- **Dhaou Ghoul** "L'apport du traitement des *tokens* arabes dans la traduction des occurrences de «lamā».". Rencontres Jeunes Chercheurs (RJC), Université Sorbonne Nouvelle, Paris, France, 9-10 Juin 2016. (<https://calenda.org/369948?file=1>)

### **Thèses et mémoires :**

- **Dhaou Ghoul**, "Classifications et grammaires des invariants lexicaux arabes en prévision d'un traitement informatique de cette langue. Construction d'un modèle théorique de l'arabe : la grammaire des invariants lexicaux temporels". Rapport de thèse, université Paris-Sorbonne, 2016. ([http://www.mmsh.univ-aix.fr/program/Documents/GHOUL\\_Dhaou\\_2016\\_These.pdf](http://www.mmsh.univ-aix.fr/program/Documents/GHOUL_Dhaou_2016_These.pdf))
- **Dhaou Ghoul**, "Outils génériques pour l'étiquetage morphosyntaxique de la langue arabe : segmentation et corpus d'entraînement. Rapport de master, université Grenoble Alpes, 2011 (<https://dumas.ccsd.cnrs.fr/dumas-00631517/document>).