



## AVIS DE SOUTENANCE DE THESE

*Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que*

**Mr AKALLOUCH Oussama**

Soutiendra : le Samedi 20/12/2025 à 15H00

Lieu : *Centre des Etudes Doctorales - USMBA – Amphi 2*

*Une thèse intitulée :*

**Modernizing Amazigh Language Processing: Artificial Intelligence-Driven Computational Methods and Technologies**

*En vue d'obtenir le Doctorat*

*FD : Sciences et Technologies de l'Information et de la Communication*

*Spécialité : Intelligence Artificielle et Big Data*

*Devant le jury composé comme suit :*

Nom et prénom	Etablissement	Grade	Qualité
LAMRINI Mohamed	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Président
El YADARI Mourad	Ecole Nationale Supérieure d'Arts et Métiers, Rabat	PES	Rapporteur
BENKAOUZ Yahya	Faculté des Sciences, Rabat	MCH	Rapporteur
BOUHOUTE Afaf	Faculté des Sciences Dhar EL Mahraz, Fès	MCH	Rapporteur
HRAOUI Said	Ecole Nationale des Sciences Appliquées, USMBA, Fès	MCH	Examinateur
ZINEDINE Ahmed	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Examinateur
FARDOUSSE Khalid	Faculté Chariaa, Fès	MCH	Directeur de thèse
BERRADA Ismail	Université Mohammed VI Polytechnique, Ben Guérir	MCH	Co-directeur de thèse



## Résumé :

La langue amazighe, parlée par des millions de personnes en Afrique du Nord et dans la diaspora, constitue une famille linguistique riche sur le plan morphologique mais largement sous-représentée sur le plan technologique. En raison de la complexité de sa morphologie, de sa variation dialectale, de la coexistence de plusieurs systèmes d'écriture (notamment le tifinagh) et du manque de ressources numériques, le développement d'outils informatiques pour l'amazighe représente un défi unique.

Cette thèse, intitulée "Modernisation du traitement de la langue amazighe : méthodes et technologies computationnelles pilotées par l'intelligence artificielle", propose une étude approfondie visant à faire progresser les technologies langagières amazighes à travers des approches modernes basées sur l'intelligence artificielle.

Le travail repose sur trois contributions principales. Tout d'abord, une revue complète des avancées récentes dans les principaux domaines du traitement automatique de la langue amazighe — traitement du langage naturel, reconnaissance vocale, reconnaissance optique de caractères (OCR) et traduction automatique — met en lumière l'évolution du domaine, notamment la transition des systèmes à base de règles vers des architectures neuronales. Cette analyse révèle des progrès inégaux selon les domaines et souligne l'importance de ces technologies pour la préservation culturelle et l'inclusion numérique.

Ensuite, la thèse explore l'apprentissage en contexte (in-context learning) pour la traduction automatique dans un contexte de faible disponibilité de données, en prenant le tarifit — une variété amazighe majeure — comme étude de cas. À travers une évaluation systématique de modèles de langage de grande taille, l'étude démontre que les techniques d'apprentissage par démonstration (few-shot learning) peuvent produire des résultats compétitifs, même pour des langues morphologiquement complexes et peu dotées. L'analyse linguistique des erreurs apporte des pistes pour l'adaptation future des modèles.

Enfin, la thèse propose TifinNet, une nouvelle architecture hybride CNN-Transformer conçue pour la reconnaissance manuscrite du tifinagh. En combinant des couches convolutives sensibles à la géométrie avec des mécanismes d'attention multi-échelle, le système atteint des performances de pointe sur des tâches complexes de reconnaissance de caractères et de mots. L'architecture répond aux défis liés à la similarité visuelle des caractères, aux variations d'orientation et à la rareté des données d'entraînement à travers des convolutions orientées et une stratégie d'apprentissage hybride.

Les contributions démontrent la pertinence et la nécessité de solutions basées sur l'intelligence artificielle pour la modernisation et la numérisation des langues marginalisées. La thèse offre à la fois des fondements théoriques et des outils pratiques, contribuant à l'avancement du traitement computationnel des langues amazighes et à la promotion d'une technologie linguistique équitable.

## Mots clés :

Linguistique computationnelle amazighe ; traitement des langues peu dotées technologies neuronales du langage ; traduction automatique ; OCR tifinagh ; préservation numérique des langues



## TMODERNIZING AMAZIGH LANGUAGE PROCESSING: ARTIFICIAL INTELLIGENCE-DRIVEN COMPUTATIONAL METHODS AND TECHNOLOGIES

### **Abstract :**

The Amazigh language, spoken by millions across North Africa and the diaspora, represents a linguistically rich yet technologically underserved language family. Due to its complex morphology, dialectal variation, multiple scripts (notably Tifinagh), and limited digital resources, developing computational tools for Amazigh poses unique challenges. This thesis, titled Modernizing Amazigh Language Processing: Artificial Intelligence-Driven Computational Methods and Technologies, presents a multifaceted investigation into advancing language technologies for Amazigh through the application of modern artificial intelligence techniques.

The work is structured around three core contributions. First, a comprehensive survey of recent advances across key Amazigh processing domains—natural language processing, speech recognition, optical character recognition, and machine translation—highlights the evolution of the field, with particular emphasis on the shift from rule-based to neural architectures. This analysis underscores disparities in progress and resource availability, and it contextualizes the role of technology in cultural preservation and digital inclusion.

Second, the thesis explores in-context learning for low-resource machine translation, using Tarifit—a major Amazigh variety—as a case study. By evaluating large language models through systematic experimentation, the study demonstrates that few-shot learning techniques can yield competitive results, even in morphologically complex, low-resource settings. The work establishes translation baselines and provides a linguistic error analysis that informs future models adaptation strategies.

Third, the thesis introduces TifinNet, a novel CNN-Transformer hybrid architecture tailored for handwritten Tifinagh script recognition. By integrating geometric-aware convolutional layers with attention-based sequence modeling, the system achieves state-of-the-art performance on challenging character-level and word-level recognition tasks. The architecture addresses core issues such as character similarity, rotational variance, and training data scarcity through oriented convolutions and a hybrid training objective.

Together, these contributions demonstrate the viability and necessity of AI-driven solutions for revitalizing and digitizing marginalized languages. The thesis offers both theoretical insights and practical tools, setting a foundation for future research in computational Amazigh linguistics and the broader effort toward equitable language technology.

### **Keywords:**

Amazigh computational linguistics; low-resource language processing; neural language technologies; machine translation; Tifinagh OCR; digital language preservation