



FICHE PROJET DE THESE pour ANNEE 2021-2022

Discipline du Doctorat	4200022 - Sciences du Langage – Linguistique
Mention du Doctorat	Sciences du langage : linguistique et phonétique générales – 7°
Domaine scientifique principal	Linguistique romane
Domaines scientifiques secondaires	-
Entités de rattachement	UMR LISA - ICPP
Direction de la thèse	Estelle RETALI-MEDORI, Dr-HDR 06 10 16 09 97 - medori_e@univ-corse.fr Pr Filipponio Lorenzo - Université Humboldt de Berlin lorenzo.filipponio@hu-berlin.de - 00491743230325 Cotutelle envisagée
Collaborations extérieures éventuelles envisagées	Université Humboldt de Berlin
Type de financement visé	Contrat doctoral
Connaissances et compétences requises chez l'étudiant	Master Sciences du Langage, Master Lettres, Master Langues, Master Langue et Culture Corses
Titre de la thèse	La syntaxe du corse et du gallurien en contexte roman : analyse d'un corpus dialectal
Abstract 1	Cette thèse aura pour objet d'étudier la syntaxe et la morphosyntaxe du corse et du gallurien en se fondant sur des corpus dialectaux oraux et écrits. L'objet est d'apporter un éclairage sur une des dimensions les moins explorées du corse en contexte roman.
Abstract 2	La thèse s'inscrit dans le cadre du programme NALC-BDLC (axe 1, ICPP, UMR LISA). Elle contribuera à une connaissance approfondie du corse ce qui est un besoin scientifique et social (besoin d'outillage du corse). Les données et leur traitement s'insèrent seront l'objet de volets BDLC (syntaxe et morphologie) et contribueront au volet Traitement Automatique des Langues appliqué au corse en cours de développement.
Explication sur page suivante	

Explicitation du Projet de thèse

1°) Présentation des aspects scientifiques du projet de thèse

Si les études consacrées à la morphologie du corse connaissent un essor important actuellement (volet de la *Banque de Données Langue Corse* et volume projeté du *Nouvel Atlas Linguistique et ethnographique de la Corse*), la syntaxe est un domaine encore peu exploré malgré quelques publications partielles ayant trait notamment à l'expression du pronom dans les subordonnées, à l'ordre des clitiques, l'usage des auxiliaires, ou encore à l'accusatif prépositionnel. Des manuels à vocation pédagogique se sont aussi attelés à des descriptions grammaticales, sans toutefois pouvoir bénéficier d'analyses suffisamment étayées en amont. Ainsi, une description syntaxique des parlers corses serait bienvenue pour enrichir la connaissance du corse sur le plan scientifique mais également pour les applications pédagogiques qui peuvent en découler.

Après avoir fait un état de l'art, la thèse devrait s'appuyer sur un corpus de données dialectales constitué des données de l'*Atlante Linguistico Etnografico Italiano della Corsica*, du corpus d'ethnotextes de la *Banque de Données Langue Corse* en lien avec les analyses réalisées actuellement dans le cadre du *Traitement Automatique des Langues* appliqué au corse, ainsi que sur un corpus spécifique, généré dans le cadre du doctorat, à partir d'un questionnaire conçu à dessein. En effet, si une part des corpus exploités par les auteurs qui se sont penchés sur la syntaxe du corse résulte d'enquêtes de terrain réalisées sous la forme d'entretiens semi-dirigés ou de liste de phrases construites afin d'identifier les éléments à étudier, une bonne part des travaux repose sur des corpus littéraires contemporains et ne permettent pas toujours d'identifier des processus langagiers spontanés ou de désambiguïser les énoncés considérés. Quant aux manuels scolaires, ils sont régulièrement construits sur la méthode de l'informant, ce qui ne permet pas d'objectiver les données traitées et induit des biais. Toutefois, les données écrites méritent d'être examinées, particulièrement dans des textes non littéraires afin de se rapprocher des productions spontanées permettant l'émergence de phénomènes plus amples.

Parmi les objectifs de recherches sur les structures syntaxiques et morphosyntaxiques du corse et du gallurien, qui devront s'inscrire dans un cadre roman et plus spécifiquement italo-roman (perspective comparatiste), il conviendrait d'approfondir les questions déjà abordées par ailleurs pour le corse ou pour d'autres variétés romanes. Ainsi, la question du corse en tant que langue présentant les caractères d'une langue *pro-drop* partiel nécessite l'analyse d'un corpus important afin d'établir les règles d'apparition du pronom dans les subordonnées ou comme anaphore, et l'analogie du corse avec des variétés romanes avec lesquelles il a été en contact (particulièrement le sarde et le toscan) peut apporter un éclairage stratigraphique sur cette question. Concernant l'accusatif prépositionnel, si certains contextes d'emploi de celui-ci ont été convenablement cernés, il convient d'en étudier plus finement les contextes notamment en lien avec les partitifs. Par ailleurs, l'ordre des clitiques, bien qu'ayant fait l'objet de quelques études, nécessite d'être mis en perspective avec d'autres variétés romanes dans une perspective diachronique et diatopique (l'ordre Objet Direct – Object Indirect est-il généralisé ? à quelle strate romane renvoie-t-il ? etc.). Également, si de l'emploi de l'auxiliaire a été abordé dans des travaux comparatistes (corse-anglais), il conviendrait éventuellement de contextualiser la question dans le cadre roman et italo-roman afin de déterminer les conditions qui déterminent le choix des locuteurs mais aussi d'établir les modalités de l'accord du participe.

D'autres aspects de l'analyse syntaxique et morphosyntaxique pourront être développés, notamment en termes de corpus car le cadre déterminé (données issues de corpus dialectologiques) pourra prendre en considération des éléments de la documentation écrite passée ou de productions orales et écrites contemporaines afin d'établir l'impact de l'interférence avec le français en contexte diglossique.

2°) Présentation des enjeux de la thèse

Par son objet (le corse) et par son approche (Sciences du Langage), cette thèse s'avère être en adéquation avec la politique scientifique de l'UCPP. Elle s'inscrit dans les développements du programme scientifique du *Nouvel Atlas Linguistique et ethnographique de la Corse* (NALC) – *Banque de Données Langue Corse* (BDLC), programme de recherches rattaché à l'Université depuis son ouverture en 1981 et actuellement réalisé au sein de l'UMR 6240 LISA. Ce doctorat contribuera à la réalisation d'un volume d'atlas linguistique (NALC) consacré à la morphologie et la syntaxe du corse dans le cadre d'une collaboration entamée sur ces thématiques avec l'Université Humboldt de Berlin.

Les données collectées par le biais d'enquêtes en complément de celles qui préexistent à la constitution du corpus, seront versées dans un volet de la BDLC (<http://bldc.univ-corse.fr>) permettant un accès à la communauté scientifique comme au grand public. Les données seront également traitées et feront l'objet de modules d'analyses – en sus de l'accès aux données brutes – grâce à un module spécifique de la BDLC et contribueront également à la production de l'outillage numérique de la langue en interaction avec le développement du *Traitement Automatique des Langues* appliqué au corse en plein essor dans la base de données (<http://bldc.univ-corse.fr/tal/>).

La méthodologie de constitution du corpus ainsi que les analyses scientifiques qui en découleront constitueront également un socle important pour la production d'outils pédagogiques. La mise à disposition du public des données et la médiation qui peut découler des travaux scientifiques pourront contribuer à la réappropriation par les habitants de la Corse, en lien avec la Gallura, des variétés menacées de disparition. Par sa nature et son objet, cette thèse répond en effet à une demande sociale manifestée en Corse qui est celle de la réappropriation et de la transmission du patrimoine linguistique de l'île.

Ce projet doctorat, qui vise à recueillir une partie du patrimoine immatériel de la Corse et de la Gallura et de le restituer par divers moyens (cf. les recommandations de l'UNESCO en la matière), entre donc en adéquation totale avec les orientations de recherche en SHS de l'UMR LISA et tout particulièrement de l'équipe ICPP.