

## APPEL À PROJETS RECHERCHE 2024

### FÉDÉRATION DE RECHERCHE AGORANTIC « CULTURE, PATRIMOINES, SOCIÉTÉS NUMÉRIQUES »

**Attention** : 5 pages [agorantic@univ-avignon.fr](mailto:agorantic@univ-avignon.fr) avant le 24 novembre 2023.

<b>Titre</b>	Étiquetage lexico-grammatical du castillan médiéval
<b>Acronyme</b>	e-CaM
<b>Nom des porteurs</b>	Olivier BRISVILLE-FERTIN (ENS de Lyon, CIHAM-UMR 5648) & Matthias GILLE LEVENSON (École nationale des Chartes-PSL, Centre Jean Mabillon-EA 3624)
<b>Gestionnaire</b>	Marie-José TORTOSA <a href="mailto:marie-jose.tortosa@univ-avignon.fr">marie-jose.tortosa@univ-avignon.fr</a>
<b>Laboratoires des membres associés</b>	EA 7412-CEIIBA (U. Toulouse JJ) ; EA 4385-LÉA (U. Paris 8) ; EA 4083-CLEA (Sorbonne U.) ; EA 4162-CeRLA (U. Lyon 2) ; ALMAnaCH, Inria ; Université de Lausanne ; Université de Genève ; Universidad Complutense de Madrid ; Universidad Autónoma de Madrid ; College Holy Cross de Worcester (MA).
<b>Budget demandé</b>	8 000 €
<b>Résumé</b>	Le projet e-CaM (« Étiquetage lexico-grammatical du castillan médiéval ») propose d'améliorer la qualité de l'annotation – lemmes, parties du discours et morphologie – des textes castillans médiévaux et de faciliter l'usage de méthodes de la philologie computationnelle pour la médiévistique hispanique. Il vise à la construction d'un corpus, représentatif de la production littéraire et documentaire des XIII <sup>e</sup> -XV <sup>e</sup> siècles, et d'un nouveau manuel d'annotation, selon des normes communément établies et précisément documentées, pour produire <i>in fine</i> les données nécessaires à l'entraînement de nouveaux modèles d'étiquetage. Le projet est constitué de spécialistes de différents domaines (linguistique, littérature médiévale, humanités numériques) qui collaboreront à ces tâches.
<b>Max. 1 000 caractères espaces comprises</b>	

#### 1. Contexte, positionnement, objectifs

L'annotation lexico-grammaticale des textes est fondamentale pour un ensemble de tâches de traitement linguistique et textuel. Pour le castillan médiéval, malgré de grandes bases de données de textes lemmatisés ([CDH](#), [OSTA](#)<sup>1</sup>) ou non lemmatisés ([CORDE](#)), il n'existe aucun outil d'annotation libre, documenté et, surtout, à jour pour analyser un texte. Le seul outil disponible est FreeLing ; sa dernière mise à jour pour le castillan médiéval date de 2012<sup>2</sup>. De son évaluation en 2021<sup>3</sup>, il ressort que l'outil ne correspond pas aux standards actuels de qualité d'annotation (entre 91 et 94 % de précision dans la reconnaissance des lemmes, des PoS et de la morphologie) ; du point de vue qualitatif, il a été mis en lumière les faiblesses de FreeLing dans la désambiguïsation des formes homographes en contexte et la forte sensibilité de l'outil à la variation graphique. Par ailleurs, son jeu d'étiquettes, [EAGLES](#), présente des défauts et se limite à une identification morphosyntaxique, à la différence de jeux plus récents comme [Universal Dependencies](#) (UD).

Pour autant, l'objectif à terme du projet e-CaM est de fournir à la communauté scientifique un modèle d'annotation efficace et actualisé, afin de faciliter les études en philologie computationnelle pour le castillan médiéval (alignement, collation), comme en lexicométrie. Elle pourra, de même, favoriser les études linguistiques diachroniques si Universal Dependencies était adopté.

Pour ce faire, les objectifs de cette première phase exploratoire, pour laquelle nous soumettons cette candidature à Agorantic, sont triples : une réflexion interdisciplinaire sur les jeux d'étiquettes afin d'améliorer l'existant ; la détermination de normes d'annotation réfléchies et documentées sous forme de manuel devant permettre, ultérieurement, de produire des données d'entraînement et des modèles efficaces et en adéquation avec les standards actuels ; enfin, la réalisation d'une première petite campagne d'annotation (le travail consistant en réalité en de la

<sup>1</sup> F. Gago Jover & F. Javier Pueyo Mena. « El Old Spanish Textual Archive, Diseño y Desarrollo de Un Corpus de Textos Medievales: Lematización y Etiquetado Gramatical ». In : *Scriptum digital. Revista de corpus diacronics i edició digital en Llengües iberoromàniques* 7 (2018), p. 25-35. url : [https://www.raco.cat/index.php/\\_scriptumdigital/article/view/343462](https://www.raco.cat/index.php/_scriptumdigital/article/view/343462). *Eid.* « El Old Spanish Textual Archive, Diseño y Desarrollo de Un Corpus de Textos Medievales: El Corpus Textual. » In : *Cuadernos del Instituto de Historia de la Lengua* 11 (2018), p. 165-209.

<sup>2</sup> C. Sánchez Marco. « Tracing the Development of Spanish Participial Constructions: An Empirical Study of Semantic Change ». Barcelone : Universitat Pompeu Fabra, 2012. url : <https://www.tdx.cat/bitstream/handle/10803/97044/tcsm.pdf?sequence=1>.

<sup>3</sup> M. Gille Levenson, O. Brisville-Fertin, M. Díez Yáñez & S. Gabay. « Construcción de Un Corpus de Evaluación de La Anotación Léxico-Gramatical Del Castellano Medieval (Siglos 13-15) ». In : *V Congreso de La Sociedad Internacional de Humanidades Digitales Hispánicas*. Santiago de Compostela, 2021. url : <https://hal.science/hal-04101415v1/file/slides.pdf>.

correction d'annotations produites par FreeLing) afin de calibrer notre manuel d'annotation et de remédier les problèmes pratiques ou théoriques qui se présenteront aux vacataires.

L'objectif à long terme d'e-CaM sera donc la production d'un corpus intégralement corrigé à la main<sup>4</sup> et l'entraînement de modèles d'annotation lexico-grammaticale. C'est la préparation méthodologique, scientifique et technique de cette campagne d'annotation qui est visée dans notre candidature à un financement Agorantic: est cherchée la normalisation de pratiques finement documentées au sein d'une communauté multidisciplinaire et internationale afin de faire émerger une large communauté d'utilisateurs et qui permette la compatibilité des jeux de données et ainsi, à terme, la constitution de grands corpus pour l'entraînement de modèles couvrant les évolutions du castillan de la période médiévale à nos jours.

## 2. Questionnement scientifique

Les avancées substantielles de l'étiquetage pour des états de langue anciens<sup>5</sup> rendent compte de potentialités certaines rendues possibles pour les Humanités Numériques. Collation, analyses sémantique, lexicale ou lexicométrique sont quelques-unes des possibilités offertes à un corpus annoté :

- du point de vue ecdotique, amélioration de la qualité de l'alignement de la collation automatisée et de la typologisation des variantes textuelles<sup>6</sup> ;
- classification d'auteurs ou vérification d'attribution, notamment avec des documents de diverses *scripta*<sup>7</sup> ;
- pour la classification sémantique des langues à morphologie riche, le lemme apporte plus aux classificateurs que les formes<sup>8</sup> ;
- facilitation de la mise en place de moteurs de recherche tels que Blacklab ou NoSketchEngine.

Nos principaux questionnements découlent d'un certain retard pris par le champ de la médiévistique médiévale hispanique par rapport, par exemple, au domaine français<sup>9</sup>. En effet, malgré ses qualités, l'outil FreeLing se fonde sur des technologies moins utilisées aujourd'hui, comme les modèles de Markov cachés (*Hidden Markov Models*, HMM) et le *relaxation labelling*<sup>10</sup> qui ont tendance à être délaissées au profit de méthodes d'apprentissage profond. Son évaluation a également mis en avant certaines limites quantitatives et qualitatives – la première étant la désambiguïsation des formes homographes – et, partant, des pistes d'améliorations. De plus, le jeu d'étiquettes de FreeLing, EAGLES, ne permet pas une catégorisation fine de certains phénomènes médiévaux (par exemple le plus-que-parfait de l'indicatif synthétique ou les pronoms adverbiaux *ý* et *ende*) et diffère de la taxonomie académique actuelle (la conjonction subordonnante *cuando* est analysée par la RAE comme un adverbe pronominal relatif). Il devient donc nécessaire aujourd'hui d'interroger ce jeu, entre linguistes, pour proposer des améliorations reposant sur des fondements scientifiques solides, mais aussi à le mettre en perspective, notamment grâce aux experts en Humanités Numériques de l'équipe, pour envisager les bénéfices ou les inconvénients d'adopter un autre jeu d'étiquettes plus récent, comme *Universal Dependencies*, utilisé par une plus grande communauté et permettant notamment d'intégrer les relations syntaxiques. Ces réflexions doivent s'effectuer en décloisonnant les disciplines et les domaines linguistiques d'expertise, notamment par le concours de spécialistes latinistes et francisants.

Nos questionnements sont donc aussi bien linguistiques que techniques et épistémologiques et posent une série de problématiques :

<sup>4</sup> Corpus dit « *gold* » en anglais. Voir M. Filannino & M. Di Bari. « Gold Standard vs. Silver Standard: The Case of Dependency Parsing for Italian ». In *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-It 2015 : 3-4 December 2015, Trento*, éd. par C. Bosco, S. Tonelli & M. Zanzotto. *Collana Dell'Associazione Italiana Di Linguistica Computazionale*. Torino: Accademia University Press, 2016. <http://books.openedition.org/aaccademia/1475>.

<sup>5</sup> Voir par exemple : <https://github.com/sigtyp/ST2024>.

<sup>6</sup> J.-B. Camps, L. Ing & Elena Spadini. « Collating medieval vernacular texts: Aligning witnesses, classifying variants ». In : *DH2019 Digital Humanities Conference*. DH2019: Complexities. Utrecht, 2019. url : <https://dh-abstracts.library.cmu.edu/works/10074>. M. Gille Levenson. « TeiCollator: A TEI to TEI Workflow ». In : *TEI2022 : Text as Data*. Newcastle, United Kingdom, sept. 2022. url : <https://hal.science/hal-03834575>.

<sup>7</sup> F. Cafiero & J.-B. Camps. « Why Molière most likely did write his plays ». *Science Advances* 5.11 (nov. 2019), eaax5489. doi : [10.1126/sciadv.aax5489](https://doi.org/10.1126/sciadv.aax5489). url : <https://shs.hal.science/halshs-02426621>. T. Clérice & A. Glaise. « Twenty-One\* Pseudo-Chrysostoms and more: authorship verification in the patristic world ». In : *Computational Humanities Research Conference 2023*. Proceedings of the Computational Humanities Research Conference 2022. Paris, France, déc. 2023. url : <https://inria.hal.science/hal-04211176>. J.-B. Camps, T. Clérice & Ariane Pinche. « Noisy medieval data, from digitized manuscript to stylometric analysis : Evaluating Paul Meyer's hagiographic hypothesis ». In : *Digital Scholarship in the Humanities* 36.Supplement 2 (2021), p. 1149-1171.

<sup>8</sup> T. Clérice. « Detecting sexual content at the sentence level in first millennium Latin texts ». working paper or preprint. Sept. 2023. url : <https://inria.hal.science/hal-04214375>.

<sup>9</sup> F. Duval, L. Ing, J.-B. Camps, N. Kanaoka, A. Pinche, et al. « Lemmatisation de l'ancien français : Présentation du modèle et des outils de l'École des chartes ». In : *XXXe Congrès International de Linguistique et de Philologie Romanes*, Société de linguistique romane, 2022, La Laguna, Tenerife, Espagne. pp.1001-1012, [10.46277/SLR.18.2023.1001-1012](https://doi.org/10.46277/SLR.18.2023.1001-1012). Jean-Baptiste Camps, Thibault Clérice, Naomi Kanaoka, Ariane Pinche, Frédéric Duval, et al. « Corpus and Models for Lemmatisation and POS-tagging of Old French ». 2021. [halshs-03353125](https://halshs-03353125)

<sup>10</sup> L. Padró. « Analizadores Multilingües en FreeLing ». *Linguamática* 3.2 (2011), p. 13-20. issn : 1647-0818. url : <https://www.linguamatica.com/index.php/linguamatica/article/view/115>.

— Comment établir un corpus représentatif du castillan médiéval ouvert et permettant divers entraînement de modèles d'annotation ?

— Quel jeu d'étiquettes présente aujourd'hui le plus d'intérêts pour la recherche actuelle et à venir ? Comment rendre l'outil compatible avec d'autres corpus linguistiques annotés ? Peut-on adopter un standard pour faciliter cette interopérabilité entre langues et envisager à plus ou moins long terme une prise en compte des relations syntaxiques ?

— Comment améliorer la qualité de l'outil d'annotation ? Peut-on envisager une méthode mixte, qui fasse usage à la fois de méthodes d'apprentissage supervisé (avec l'outil Pie, par exemple) tout en réutilisant une partie des données produites par l'équipe de Freeing, qui se présentent en particulier sous la forme d'une liste conséquente de formes annotées ?

### 3. Méthodologie

Le [CORDE](#) (Corpus Diacrónico Del Español) sera mis à contribution pour récupérer automatiquement un corpus massif, qui sera réduit et équilibré : le corpus cible comptera entre 1,5 et 2 millions de mots, extension comparable aux corpus latin ([LASLA](#)) et d'ancien français ([OF3C](#)). Ce corpus sera **annoté** à l'aide de l'outil d'annotation Freeing, puis intégralement **corrigé** par une équipe constituée d'une dizaine de vacataires, supervisés par les membres du projet. Le financement demandé à Agorantic permettra plusieurs orientations de travail :

**A. La création du corpus** (déjà ébauché [[LIEN](#)]), mais susceptible d'être largement remanié) occupera les membres, notamment les philologues médiévistes, dans les premiers mois du projet. Les données textuelles seront récupérées avec leurs métadonnées afin de pouvoir facilement équilibrer le corpus (du point de vue générique comme historique). Nous visons au moins un millier d'œuvres distinctes. La datation proposée par CORDE étant problématique – car les textes sont datés selon la date estimée de création de l'œuvre et non pas de production du manuscrit<sup>11</sup> – un travail de vérification et de correction de la datation des œuvres est envisagé et sera nécessaire. Du point de vue générique, nous visons l'équilibre entre œuvres littéraires (droit et codification inclus) et documents de la pratique. Les œuvres seront classées en genres et sous-genres qui suivront autant que possible la [catégorisation](#) de la Real Academia Española. Du point de vue des normes de transcription, le corpus CORDE est particulièrement intéressant, car il intègre de très nombreuses œuvres produites par différents éditeurs et éditrices aux normes d'édition changeantes. **Le corpus reflétera donc à la fois la variabilité graphique** de la langue castillane médiévale comme la **variabilité des normes d'édition** qui sont en vigueur dans la philologie castillane médiévale actuelle : le modèle ainsi entraîné pourra donc être efficace avec de nombreuses normes de transcription.

**B. La deuxième tâche** consistera en une **réflexion sur les normes d'annotation** et les avantages et inconvénients entre jeux d'étiquettes possibles (EAGLES ou *Universal Dependencies*). Elle portera de même sur l'analyse des faiblesses de Freeing, complétant ainsi le travail déjà amorcé<sup>12</sup> (exemples : normalisation des ramistes, gestion des contractions). Ce faisant, nous pourrons **orienter le travail de correction** pour prioriser des éléments-clés ou jugés prioritaires : la gestion de l'homographie sera un des enjeux principaux des corrections. Cette réflexion débouchera sur la **création d'un guide d'annotation**, qui sera amendé à l'issue de la première phase de correction test (tâche **C**). Une liste de contrôle de lemmes et de parties du discours et morphologie sera constituée à partir du travail des membres de Freeing et du projet OSTA en ce qui concerne les lemmes, et en fonction du jeu d'étiquettes choisi en ce qui concerne l'annotation grammaticale et morphologique.

Pendant les jours d'atelier, nous proposerons une formation aux pratiques et enjeux de l'étiquetage lexico-grammatical du castillan médiéval (normes, utilisations des outils et des bases de données annotées) à destination des vacataires, qui seront ouvertes au public.

**C. Une troisième tâche**, qui sera réalisée entre les deux ateliers, consistera en une **campagne de correction-test** d'une portion du corpus (environ 120 heures). Elle permettra une première prise en main de l'outil Pyrrha<sup>13</sup>, qui sera utilisé par les correcteurs et par les superviseurs, et permettra de tester un outil de correction en contexte qui a commencé à être développé par M. Gille Levenson [[LIEN](#)]. Cette série permettra de mettre à l'épreuve la méthodologie choisie, et de faire apparaître des problèmes éventuels.

Ces trois tâches seront mises en commun et leurs résultats discutés et synthétisés lors des deux ateliers de travail prévus.

<sup>11</sup> J. Rodríguez Molina & Á. Octavio. de Toledo y Huerta. «La imprescindible distinción entre texto y testimonio: el CORDE y los criterios de fiabilidad lingüística». *Scriptum digital. Revista de corpus diacrò nics i edició digital en Llengües iberoromàniques* 6 (2017), p. 5-68, <https://raco.cat/index.php/scriptumdigital/article/view/329258>

<sup>12</sup> M. Gille Levenson, O. Brisville-Fertin, M. Díez Yáñez & S. Gabay. « Construcción de un corpus de evaluación... ». M. Gille Levenson. « Le *Regimiento de Los Príncipes* et Sa Glose : Étude et Édition numérique de la partie sur le gouvernement de la cité en temps de guerre (III, 3). » Thèse de doct. École Normale Supérieure de Lyon, 2023.

<sup>13</sup> T. Clérice, V. Jolivet, J. Pilla. « Building infrastructure for annotating medieval, classical and pre-orthographic languages: the Pyrrha ecosystem ». *Digital Humanities 2022* (2022), Tokyo. [hal-03606756](https://hal-03606756)

Le premier atelier (mars-avril) permettra de lancer le projet ; il permettra la constitution du corpus et de l'ajuster éventuellement à la marge. Il sera essentiellement consacré à la discussion sur le jeu d'étiquette d'annotations grammaticales et morphologiques, et abordera aussi la question de la priorisation des corrections.

L'objet principal du second atelier (octobre-novembre) sera la rédaction d'un manuel d'annotation du castillan médiéval qui soit à la fois descriptif (documentation de nos pratiques) et prescriptif (indication de la marche à suivre pour augmenter le corpus ou annoter de nouvelles données textuelles). L'atelier permettra aussi de discuter des difficultés des transcripteurs et des problèmes posés par les choix collectifs, afin d'ajuster et d'amender le guide d'annotation

#### 4. Résultats attendus et caractère innovant de la recherche

1. Des décisions documentées sur les jeux d'étiquette et les référentiels de lemmes, sous la forme d'un **manuel d'annotation** qui vienne répondre aux limites de Freeling, voire d'EAGLES.

2. Un **corpus intégralement corrigé à la main et vérifié par des spécialistes** qui permettra l'entraînement d'un modèle d'annotation efficace, l'objectif étant à terme de dépasser Freeling (le corpus d'évaluation produit en 2021 sera mis à contribution pour comparer les deux outils). Ce corpus sera proposé sous différents formats : TSV mais aussi XML-TEI, ce qui rendra possible sa réutilisation pour d'autres projets de recherche.

3. Un **modèle d'annotation** efficace entraîné à partir des données corrigées.

4. Des **productions scientifiques** (articles, *datapapers* en particulier, et communications).

C'est la première production, que nous estimons fondamentale, qui sera permise spécifiquement par Agorantic, par le biais du financement de deux ateliers de travail en présentiel et d'une première salve de corrections, ainsi qu'une première communication en congrès qui permettra de présenter le projet.

Notre projet global s'inscrit pleinement dans le paradigme de la **science ouverte** : ouverture des données, des modèles et de la documentation. Il vise à donner accès à la fois à un outil d'annotation (le modèle), mais aussi à un **corpus structuré** et à une **méthodologie finement documentée**. L'intégralité des productions du projet (techniques, linguistiques, scientifiques) seront publiées sur des plateformes : [Nakala](#) ou [recherche.data.gouv.fr](#) pour les données ainsi que [Gitlab](#) et Zenodo pour les scripts ; [HAL](#) pour la version *preprint* du *datapaper* prévu et du manuel d'annotation.

Le manuel d'annotation sera donc la principale réalisation permise par le financement Agorantic et sera rapidement publié sur HAL.

#### 5. Dimension interdisciplinaire et inscription dans les axes 1 et 5 de la FR Agorantic

Le projet rassemblera des spécialistes de linguistique (diachronique, systémique et lexicologique), d'Humanités Numériques et des médiévistes, travaillant aussi bien sur les aires hispanophones que francophones et latines. La plupart des membres présentent d'ailleurs des profils interdisciplinaires. Ils œuvreront ensemble à la réflexion sur les jeux d'étiquettes lexico-grammaticales et à leurs enjeux, à la constitution d'un corpus, à la production d'outils d'aide à la correction et à l'élaboration d'un manuel d'annotation devant, à terme, permettre la production de données d'entraînement et de modèles d'étiquetage.

Le projet e-CaM rejoint les objectifs de l'axe 5 (Structuration et exploitation de corpus) de la FR Agorantic, en cela qu'il veut adapter une méthodologie développée pour d'autres états de langue anciens, notamment le français, pour élaborer un modèle aujourd'hui manquant d'annotation lexico-grammaticales de textes médiévaux hispaniques, en vue d'en faciliter l'exploitation philologique, linguistique et historique. Le travail de réflexion qui sera mené à bien dans le cadre de la candidature Agorantic s'avérera primordial pour la phase suivante de construction des corpus et d'entraînement de modèles d'apprentissage supervisé.

Ces enjeux méthodologiques et interdisciplinaires rejoignent donc également les objectifs de l'axe 1 (Méthodologies et Interdisciplinarité), lequel promeut la collaboration entre SHS et SEE, la mutualisation d'outils et d'approches et le soutien à des phases exploratoires pour des projets de plus grande envergure.

#### 6. Partenariats

L'équipe réunit des linguistes et des médiévistes de 6 laboratoires français distincts et des universités Complutense et Autonome de Madrid, de Lausanne et de Genève. Elle compte également 5 membres qui sont spécialistes en Humanités Numériques – Ariane Pinche, CR CNRS au CIHAM, Matthias Gille Levenson, post-doctorant au Centre Mabillon, Simon Gabay, Maître assistant à l'Université de Genève, Francisco Gago, Professor au College of the Holy Cross, et Thibault Clérice, Starting Research Position à ALMAnaCH, équipe de recherche en TAL de l'INRIA.

Une demande d'aide au financement a été soumise pour l'axe Copor@tech-Sciences ouvertes du CIHAM.

Comme précisé plus bas, nous allons demander l'aide d'Huma-Num pour la plateforme de correction Pyrrha.

#### 7. Objectifs de pérennisation du projet

Cette candidature est proposée en complément d'une seconde candidature à l'appel à projets exploratoires de Biblissima+ 2024 sur le même objet (pré-candidature acceptée par le comité d'évaluation de Biblissima+ 19/11/23; dépôt du dossier le 31/01/24). Les modalités de financement et les calendriers se révèlent complémentaires, puisqu'Agorantic envisage le financement de vacations et de réunions, alors que Biblissima+ ne permet pas le financement de réunions de travail et d'ateliers: le financement demandé à Biblissima+ (20 000 €) sera intégralement consacré, en cas de succès, à des vacations de correction du corpus annoté. La double série d'ateliers prévue en 2024

est fondamentale pour notre démarche et la bonne réalisation du projet dans son ensemble. Il requiert en effet de réelles séances de travail collectif, permettant synthèses, réflexions et échanges entre disciplines et entre champs, afin de définir les meilleures solutions pour la production des livrables, à savoir le corpus *gold*, le manuel d'annotation et les modèles. Cette première phase fondamentale permettra la poursuite du projet lors d'une deuxième phase de correction du corpus, beaucoup plus ambitieuse (environ 850 heures de vacation) et d'entraînement de modèles d'annotation.

#### 8. Expression des besoins en assistance informatique :

La plateforme sous licence libre Pyrrha sera utilisée pour la correction des annotations. L'équipe ALMAnaCH (Inria) via un hébergement chez Huma-Num est en train de déployer une instance nationale pour les équipes de recherche souhaitant l'utiliser. Un hébergement d'un service d'annotation automatique pourra compléter une telle offre

9. Budget prévisionnel	Brève description	Montant
Organisation de réunions	Atelier de lancement & Atelier de consolidation et finalisation du manuel : 14 participants, 2 x 2 jours.	5 670,40 €
Vacations	120 h de vacations x 22,33 €	2 679,60 €
Budget total		8 350 €
Co financements le cas échéant	CIHAM, axe Corpor@tech (demande en 11/2023)	350 € (en cours)
<b>Budget demandé à Agorantic</b>		<b>8 000 €</b>

#### 10. Planification

e-CaM	1	2	3	4	5	6	7	8	9	10	11	12
	Phase préparatoire											
Phase 1 Agorantic				1 <sup>er</sup> atelier								
					Correction				Analyse			
										2 <sup>e</sup> atelier et manuel		

#### 11. Équipe détaillée

Porteurs : [Olivier Brisville-Fertin](#) (MC), UMR 5648–CIHAM, diachronicien et philologue médiéviste hispaniste ; [Matthias Gille Levenson](#) (post-doc.), EA 3624–Centre Jean Mabillon, spécialiste HN et philologue médiéviste hispaniste.

Membres : [Amaia Arizaleta](#) (PU), EA 7412–CEIIBA, philologue médiéviste hispaniste ; [Mónica Castillo Lluch](#) (Professeure Ordinaire), Université de Lausanne, diachronicienne et philologue médiéviste hispaniste ; [Thibault Clérice](#) (Starting Research Position), équipe ALMAnaCH, INRIA, spécialiste HN et latiniste ; [María Díez Yáñez](#) (Asistente doctora), U. Complutense de Madrid, philologue médiéviste hispaniste ; [Simon Gabay](#) (Maître assistant), U. de Genève, spécialiste HN et philologue médiéviste et moderniste francisant ; [Francisco Gago Jover](#) (Professor), College of the Holy Cross de Worcester (MA, USA), philologue médiéviste hispaniste et spécialiste HN ; [Mallorie Labrousse](#) (MC), EA 7412–CEIIBA, diachronicienne hispaniste ; [Marta López Izquierdo](#) (PU), EA 4385–LER, diachronicienne hispaniste et spécialiste HN ; [Corinne Mencé-Caster](#) (PU), EA 4083–CLEA, diachronicienne hispaniste ; [Ariane Pinche](#) (CR), CNRS-UMR 5648–CIHAM, philologue médiéviste francisante et spécialiste HN ; [Irene Salvo García](#) (chercheuse Ramón y Cajal), U. Autónoma de Madrid, philologue médiéviste hispaniste ; [Belén Villar Díaz](#) (MC), EA 4162–CeRLA, diachronicienne et lexicologue hispaniste.