

**S'écouter lire : vers une grammaire historique des pratiques de lecture**  
**(Domaine français. 13<sup>e</sup>-18<sup>e</sup> siècles)**

*English version available*

**Tableau récapitulatif des personnes impliquées dans le projet**

PARTENAIRES	NOM	PRÉNOM	POSITION ACTUELLE	RÔLE [① : numéro de tâche]	Implication PERS/MOIS
<b>EA 7345 CLESTHIA</b> [en minuscules, "Clesthia", quand rattachement secondaire pour le projet]	Badiou-Monferran	Claire	PR	Coordnatrice scientifique. / Référente Clesthia ① ② ③-⑨ ⑩ ⑪ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ ⑲ ⑳ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚	25 mois
—	Bergeron-Maguire	Myriam	MCF	③-⑨ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖	4.2 mois
—	Halévy	Olivier	MCF	③-⑨ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	4.2 mois
—	Lefevre	Florence	PR	③-⑨ ㉑ ㉒ ㉓ ㉔	4.2 mois
—	Lombardero	Emily	ATER	③-⑨ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚	5 mois
—	Maggi	Ludovica	MCF	③-⑨ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚	6 mois
Clesthia/ FIRL	Régent-Susini	Anne	PR	③-⑨ ⑫ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚	5 mois
Clesthia/ RCFC Turin	Amatuzzi*	Antonella	PR Associata	③-⑨ ⑫ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	6 mois
Clesthia/ RCFC Potsdam	Gerstenberg*	Annette	Prof. Dr.	③-⑨ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝ ㉞	6 mois
Clesthia/RCFC Neuchâtel	Skupien Dekens*	Carine	PR	③-⑨ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝ ㉞ ㉟ ㊱ ㊲	6 mois
Clesthia/ ITEM	Réach-Ngô*	Anne	MCF HDR	Référente E.C. Humanités Numériques ③-⑨ ⑩ ⑪ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ ⑲ ⑳ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛	8 mois
Clesthia /THALIM	Walter*	Richard	IG (Ing. Rech)	Gestionnaire de la Plateforme EMAN ③ ④ ⑤-⑨ ⑩ ⑪ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ ⑲ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗	10 mois
CLESTHIA	XXX	XXX	Ingénieur.e.s d'étude 1/2	IG 1 [Mois 2-13] : ⑪ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ ㉑ ㉒ IG 2 [Mois 24-36] : ㉔ ㉕ ㉖	12 mois 12 mois
CLESTHIA	XXX	XXX	Doctorant.e [Mois 7-42]	④ ⑥-⑨ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝ ㉞ ㉟	36 mois
<b>EA 4509 STIH</b> [en min., "Stih", quand rattachement 2 <sup>aire</sup> /projet]	Parussa	Gabriella	PR	Référente STIH ③-⑨ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝ ㉞	13 mois
—	Biu	Hélène	MCF	③-⑨ ⑫ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	6 mois
—	Desbois-Ientile	Adeline	MCF	③-⑨ ⑫ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	6 mois
—	Ducos	Joëlle	PR	③-⑨ ⑫ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	5 mois
—	Lejeune	Gaël	MCF	Référent TAL ③-⑨ ⑩ ⑪ ⑫ ⑬ ⑭ ⑮ ⑯ ⑰ ⑱ ⑲ ⑳ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜	18 mois
Stih / CLESTHIA	Valentini**	Andrea	MCF HDR	③-⑨ ⑫ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	6 mois
Stih / IEMYRhd	Llamas Pombo**	Elena	PR Titular	③-⑨ ⑫ ⑬ ⑭ ⑮ ㉑ ㉒ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	13 mois
	XXX	XXX	Post-Doc [Mois 18-41]	Annotation/enrichissement des corpus ④ ⑧ ⑨ ⑮ ㉔ ㉕	24 mois
<b>UMR 9193 SCALab</b>	Casalis	Séverine	PR	Référente SCALab ③-⑨ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝ ㉞ ㉟ ㊱ ㊲ ㊳	13 mois
—	Javourey-Drevet	Ludivine	MCF	③-⑨ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	8 mois
—	Mahé	Gwendoline	MCF	③-⑨ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	2 mois
—	Sparrow	Laurent	MCF	③-⑨ ㉓ ㉔ ㉕ ㉖ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	2 mois
	XXX	XXX	Post-Doc [Mois 13-28]	④ ⑦ ㉗ ㉘ ㉙ ㉚ ㉛ ㉜ ㉝	18 mois

**EA 7345 CLESTHIA** (Clesthia) : Corpus, Linguistique, Énonciation, Sens, Textes, Histoire, Interaction, Acquisition »

**EA 174 FIRL** : Formes et Idées de la Renaissance aux *Lumières* ; **RCFC** : Réseau transnational : Réseau Corpus Français Préclassique et Classique

**UMR 8132 ITEM** : Institut des Textes et Manuscrits modernes ; **UMR 7172 THALIM** : Théorie, Histoire des Arts et Littératures de la Modernité

**EA 4509 STIH** (Stih) : Sens, Texte, Informatique, Histoire ; **IEMYRhd** : *Instituto de Estudios Medievales y Renacentistas y de Humanidades Digitales*

**UMR 9193 SCALab** : Sciences cognitives et sciences affectives

\* : EC et IE participant à l'équipe projet au sein de CLESTHIA

\*\* : EC participant à l'équipe projet au sein de STIH

## I. Contexte, positionnement et objectifs de la proposition

Situé à l'articulation des sciences de la culture, des sciences du langage et de la psycholinguistique, le projet **EcoLe** (« S'ECOuter Lire ») vise à l'établissement d'une **nouvelle grammaire historique** : celle, pionnière, **des diverses pratiques de lecture (oralisée, silencieuse et déclamée)** qui, en France, ont coexisté et connu une intense période de covariation entre le début du Moyen-Âge central (13<sup>e</sup> s.) et la fin de l'Ancien régime (18<sup>e</sup> s.). Il se propose d'étudier la manière dont les écrits de cette période ont pu encoder leur mode d'appropriation préférentiel (par l'œil ou par l'oreille).

**La collaboration**, inédite, de deux équipes de linguistes **diachroniciens** avec une équipe de **psycholinguistes** (voir le consortium ci-dessus et ci-dessous) permettra de **revenir aux enjeux contemporains de la lecture** – notamment, au récent retour en force de la lecture « à haute voix », non seulement sur la scène culturelle mais aussi dans les programmes scolaires et dans les protocoles de remédiation des troubles de la lecture (e.g. dyslexies) – et de faire ainsi profiter la réflexion des acquis de ce détour par l'examen des pratiques de lecture du passé. Cette synergie, combinant les apports de la philologie, de la grammaire textuelle et de la psycholinguistique, se veut novatrice et vise à favoriser **l'émergence d'une nouvelle discipline : la psycholinguistique diachronique**.

### I.A. Hypothèses de travail – Objectifs

#### I.A.1 Rappel des hypothèses de recherche

Le projet EcoLe fait converger plusieurs hypothèses de recherche : linguistiques ; psycholinguistique ; interdisciplinaire.

#### 1. Hypothèses de linguistique diachronique

**Contexte** – Le régime d'historicité des pratiques de lecture, telles qu'elles ont été et sont encore en France et en Europe, fait aujourd'hui débat. Deux descriptions en sont proposées. **La première, évolutive**, documente la double transition ayant conduit, selon elle, de la lecture à haute voix à la lecture silencieuse (Zumthor 1987 ; Cavallo et Chartier 1997 ; Saenger 1997), et de la lecture publique, collective, à la lecture privée, individuelle (Chaytor 1945, Bouchet 2008). **La seconde, variationniste ou circonstancielle**, postule l'existence de toutes les pratiques de lecture (pour l'oreille *vs* pour l'œil) à toutes les époques – Antiquité (références dans Haug 2014) ; Moyen Âge (Coleman 1996) ; Temps modernes (Darnton 1971 ; Gilmont 2003) ; Époque contemporaine (références dans Haug 2014) –, le mode de partage des textes dépendant de la situation de communication. Pour autant, par-delà leurs divergences de vues, tous les historiens et historiennes suscités conviennent que jamais, la coexistence des trois régimes de lecture :

- (1) celui, ancestral, de la « haute voix » et de son variant, la « ruminant » (faisant se succéder trois opérations : je lis ; j'oralise ; je comprends ce que j'ai lu et oralisé)
- (2) celui, concurrent, de la lecture « silencieuse » (constitué de deux opérations : je lis ; je comprends)
- (3) celui de son avatar vocalisé, « la déclamation » (se déclinant en trois opérations : je lis ; je comprends ; et ultérieurement, j'oralise pour moi-même ou pour un auditoire ce que j'ai préalablement lu et compris en régime de lecture silencieuse)

n'a été plus intense, en France, que pendant la période conduisant du Moyen Âge tardif (14<sup>e</sup>-15<sup>e</sup> siècles) à la Révolution (fin du 18<sup>e</sup> siècle). Pour son propos, le projet EcoLe fait le choix de retenir cette période, sélectionnée pour son exemplarité, sans s'interdire des excursions en amont (13<sup>e</sup> s.) et en aval (du 19<sup>e</sup> au 21<sup>e</sup> s.). **Hypothèses** – Deux hypothèses alimentent sa réflexion. **Hypothèse<sub>1</sub>**. La première postule l'existence, historique, d'une corrélation entre la qualité formelle (*ie*, l'architecture textuelle) des textes et leur modalité de lecture (par l'œil ou par l'oreille). **À chaque architecture textuelle correspondrait/aurait correspondu un mode d'appropriation (visuel vs vocal) adapté**. Autrement dit, EcoLe formule l'hypothèse, inédite, que durant la période de covariation étudiée, les productions écrites ont pu encoder, *via* une forme de séquençage syntaxique et de structuration textuelle spécifiques, le mode de lecture auquel elles étaient préférentiellement destinées. Sur la base de sondages préliminaires (voir entre autres Badiou-Monferran 2017), il gage que les scripteurs du passé programmaient le mode d'appropriation de leurs textes, non pas tant par des prescriptions de lecture explicites que par une série de choix matériels et linguistiques, plus ou moins conscientisés, qui en facilitaient la consommation orale ou visuelle. **Hypothèse<sub>2</sub>**. La seconde hypothèse, qui justifie la collaboration de deux équipes de linguistes, l'une spécialiste de français médiéval et d'*Early Renaissance*, l'autre, de français préclassique et classique, est que, pour des raisons qu'il conviendra d'identifier, les formats textuels, du Moyen Âge tardif à la Révolution, ont évolué. La trajectoire, à nuancer selon les genres de discours, conduirait d'un **format bimodal** (“bimodality”, Coleman 1996), où, moyennant quelques adaptations de mise en page et de ponctuation, un même écrit se prêtait aux deux modalités de lecture (à voix haute ou silencieuse), à des **architectures plus nettement différenciées** : (a) l'une, congruente avec le régime de « haute voix », (i) multipliant les phrases très longues (les « périodes ») (ii) comprenant peu de ponctuation et beaucoup de connecteurs (iii) privilégiant, au sein des chaînes de référence, les marqueurs de redénomination (noms propres, SN) sur les pronoms personnels (iv) et faisant l'objet d'une mise en page compactée (sans paragraphe ni alinéa) ; (b) l'autre, destinée *a priori* à la lecture silencieuse (i) multipliant les phrases courtes et comprenant du coup beaucoup de signes de ponctuation (ii) mais mobilisant peu de connecteurs, (iii) usant davantage, dans les chaînes de référence,

des pronoms personnels que des noms propres et des SN (iv) et faisant l'objet d'une mise en page aérée (avec paragraphes et alinéas). C'est ce moment de bascule, temporellement variable suivant les genres de discours, qu'EcoLe se propose de repérer et de décrire, avec l'**outillage de la linguistique de corpus**. **Enjeux** – En adoptant un **point de vue interne** et en recourant aux **méthodes d'analyse de la linguistique textuelle**, EcoLe ne vise pas seulement à vérifier, pour les écrits dont le mode de lecture préférentiel est connu – car il est explicité ou a pu être déduit d'autres sources –, la corrélation entre format textuel et modalité de lecture ; il entend aussi reconstruire, uniquement à partir d'**indices textuels**, des **pratiques de lecture probables**. Cette démarche, **probabiliste**, n'est certes pas en mesure de produire sa propre instance de validation. Elle n'en présente pas moins l'avantage, par rapport à l'approche positiviste des historiens, de pouvoir faire émerger des hypothèses de lecture là où les témoignages externes manquent.

### 1.2 Hypothèse psycholinguistique

Le questionnement des psycholinguistes engagés dans EcoLe s'inscrit dans le prolongement de celui des linguistes diachroniciens. Mais il va plus loin. Introduisant un troisième terme, il interroge, non pas seulement la corrélation entre formats textuels et modalités de lecture, mais l'**impact de cette corrélation sur le degré de compréhension en lecture**. Son hypothèse, inédite, est qu'il existe peut-être un **lien de causalité** entre les mécanismes de compréhension en lecture et la présence *vs* absence d'une corrélation entre modalité de lecture (à voix haute *vs* silencieuse) et format (visuel *vs* vocal ; autrement dit, tabulaire *vs* linéaire) du texte lu.

### 1.3 Hypothèse interdisciplinaire

Le présupposé à l'origine de la collaboration entre deux équipes de linguistes et une équipe de psycholinguistes est que la mise en commun des résultats des unes et des autres n'est pas cosmétique : que, pour pouvoir produire ces derniers, les psycholinguistes ont besoin des linguistes et réciproquement. Il ne s'agit pas seulement d'apporter des contributions de nature différentes à l'*intelligence des conditions formelles de la lisibilité*. Le **modèle collaboratif** ici proposé va **bien au-delà de la simple pluridisciplinarité**. À supposer en effet que les textes modernes et contemporains, imprégnés d'une culture de l'écrit liée à la double révolution de la Presse et de l'Hypertexte (Vandendorpe 1999), ne puissent probablement plus programmer qu'un seul régime textuel, celui de l'appropriation visuelle – et ce, quand bien même ils feraient l'objet de lectures (contre-nature ?) à haute voix –, les expérimentations psycholinguistiques requises devront se nourrir de corpus textuels plus anciens. Réciproquement, les expérimentations des psycholinguistes viendront alimenter les modélisations des linguistes. Elles en constitueront l'instrument de validation (*vs* invalidation) empirique. On conviendra que de bons résultats aux tests de compréhension puissent alimenter, sans en garantir l'effectivité, l'hypothèse, soutenue en diachronie, d'une corrélation entre architecture textuelle et modalité de lecture. Les **quatre objectifs** présentés *infra* visent à établir les conditions de possibilités, techniques et scientifiques, de cette **entreprise transdisciplinaire**. Leur articulation est figurée dans le schéma ci-dessous :

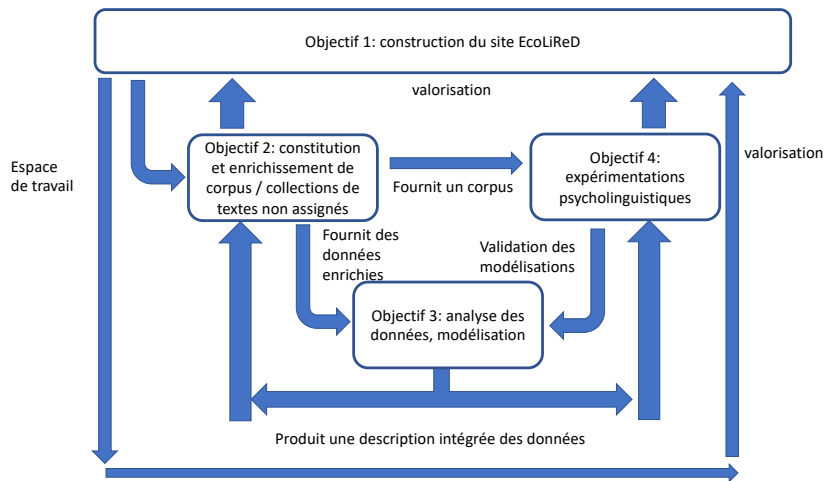


Fig. 1 : articulation des objectifs conditionnant l'élaboration d'une *Grammaire historique des pratiques de lecture*

#### I.A.2 Présentation des objectifs

##### Objectif 1 – Création d'un site collaboratif : ÉcoLiReD (« ÉCOOuter LIRE/REgarder Dire »)

Tout à la fois méthodologique, numérique et éditorial, cet objectif prélude aux objectifs scientifiques 2, 3 et 4. Il constitue, en amont, la condition de possibilité de leur réalisation et, en aval, la condition de possibilité de leur valorisation. La construction du site collaboratif EcoLiReD entend :

- ♦ constituer un outil de travail collaboratif ergonomique, facile d'utilisation pour des chercheurs non spécialistes des Humanités numériques, afin de rassembler, d'éditer et d'exploiter des données, en proposer un premier traitement textométrique, en assurant leur interopérabilité, ainsi que la valorisation des résultats de la recherche.

- ◆ élaborer un outil de consultation et d'interrogation du corpus par les divers publics visés par le projet : spécialistes de linguistique, de psycholinguistique mais aussi historiens des pratiques culturelles, sociologues, littéraires, etc. Les interfaces et les outils conçus devront également être accessibles à un plus large public, notamment les enseignants du second degré ainsi que les publics scolaire et étudiant.

Le site EcoLiReD s'appuiera pour cela sur les outils et les méthodologies proposés par la plateforme EMAN développée par Richard Walter au sein du laboratoire THALIM. Une cinquantaine de projets y participe, avec notamment un fort pôle en cours de développement en ce qui concerne la Première Modernité. Chaque projet a un site autonome sur la plateforme. Mais le site EcoLiReD pourra bénéficier de l'expérience d'autres projets sur EMAN qui portent sur le Moyen Âge et la Première Modernité.

Fondée sur le respect des standards, cette plate-forme utilise des langages ouverts et met à la disposition ses réalisations techniques auprès des communautés informatiques. Elle est conçue au départ sur une version du logiciel de bibliothèque numérique Omeka, offrant un espace de structuration des données en adéquation avec les principes FAIR et proposant une ergonomie simple d'utilisation pour les contributeurs, avec un espace de travail en ligne convivial (formulaire de saisie en WySiWyg – what you is what you get).

Tout en respectant les orientations d'Omeka, EMAN fournit un grand nombre de fonctionnalités supplémentaires qui permettent d'aller vers l'édition critique et vers l'exploitation de données. Pour l'interopérabilité de celles-ci, le site sera adossé à un entrepôt de données qui permettra un dialogue entre les corpus des équipes mais aussi avec des corpus partenaires. La plateforme EMAN est hébergée par Huma-Num qui en assure une disponibilité maximale.

EMAN présente l'avantage d'être évolutive et collaborative. Elle pourra ainsi intégrer les nouveaux besoins de visualisation et d'exploitation que le projet demandera. En particulier, elle devra permettre la double annotation des données insérées sur EcoLiReD :

- ◆ des phénomènes textuels qui nécessitent des outils de traitement linguistique, notamment des outils d'étiquetage sémantique poussé en XML-TEI, afin d'encoder le comportement des principaux observables du projet (connecteurs, ponctuation, paragraphe, etc.)
- ◆ des phénomènes typo-iconographiques, avec une granularité fine permettant d'identifier le rôle de la mise en espace des textes, la place des ornements, *marginalia*, hiérarchisation des titres, alternance des types de caractères, afin de restituer les phénomènes visuels qui interviennent dans l'opération de lecture.

Via l'utilisation originale d'Omeka comme bibliothèque numérique plaçant au cœur de la démarche d'édition le document image, il sera notamment possible de procéder à un traitement des pages numérisées, afin de réaliser un encodage des phénomènes typo-iconographiques. EMAN devra déployer des modules de liaison avec les outils d'OCR et d'HTR pour pouvoir intégrer une transcription des textes qui permette d'avoir très vite la structuration typographique et visuelle des imprimés. Il faudra aussi réaliser un certain nombre de plugins pour permettre l'intégration de ces données et leur traitement lexicométrique, en se basant sur un dialogue avec des outils existants et robustes.

## Objectif 2. – Construction et annotation de nouveaux corpus

Les équipes STIH (équipe de français médiéval et renaissant) et CLESTHIA (équipe de français préclassique et classique) entendent mettre à la disposition des communautés scientifiques de nouveaux types de corpus, non institués, inédits, transcendant les frontières traditionnelles de genres de discours et des types de textes : ceux tenant compte du **mode d'appropriation des textes, visuel et/ou vocal**.

Conformément à l'hypothèse de travail exposée *supra*, **l'équipe STIH** se fixe comme objectif de faire émerger un **corpus textuel bimodal**, à savoir : des textes pour lesquels les deux modalités de lecture à voix haute et silencieuse sont possibles. L'hypothèse qu'il s'agira de vérifier après analyse qualitative et quantitative du corpus est que, par-delà les traditionnelles variantes textuelles, la mise en page (ponctuation, usage des majuscules, espaces blancs, distribution du texte sur la page, etc.) et/ou l'iconographie peuvent varier en fonction de l'utilisation qui est programmée de chaque témoin imprimé ou manuscrit.

Suivant une approche diachronique, cette analyse visera aussi bien à déceler les marques de la langue écrite qui évoluent au cours du temps que leur rapport avec les pratiques de lecture. Une telle démarche implique la nécessité de comparer différents témoins de chaque texte, en tenant compte des variables *chronologique* (du 13<sup>e</sup> au 16<sup>e</sup> s.), *formelle* (vers / prose), de *support* (manuscrit et imprimé), de *destination* (lecteur individuel, public d'auditeurs, etc.) et de *genre textuel*.

Le corpus sera donc constitué de textes historiques, de textes littéraires (théâtre, romans), de livres de dévotion et de sermons, de textes juridiques et d'ouvrages à contenu scientifique, didactique ou encyclopédique, dont chaque spécialiste de l'équipe a sélectionné les versions les plus représentatives pour une histoire des rapports entre la matérialité de la langue écrite et l'histoire de la lecture.

L'équipe CLESTHIA mettra pour sa part à l'épreuve des textes l'hypothèse selon laquelle l'art de lire le français, sous l'Ancien Régime, s'est réalisé sous **trois formats**, adossés chacun à des architectures textuelles différenciées, codant une manière spécifique de percevoir, mémoriser et comprendre : l'un, de la **haute voix**, l'autre, de la **lecture silencieuse**, le troisième, hérité de la période précédente, **multimodal**. Elle envisage ainsi de construire un corpus modulaire, mettant en regard trois sous-ensembles :

- ◆ un sous-ensemble de textes entretenant un **rapport plus ou moins explicite avec « la haute voix »**. Suivant un modèle de catégorisation graduelle emprunté à la théorie du prototype, ce sous-ensemble contiendra, entre autres : (a) un noyau d'écrits apparentés au genre des « **cris publics** », ensemble d'actes administratifs lus en place publique par des jurés crieurs avant d'être placardés et affichant explicitement un mode de consommation par « haute et intelligible voix » ; (b) un ensemble longitudinal,

plus périphérique, de **sermons** protestants (corpus Sermo <http://sermo.unine.ch/SERMO/>) et catholiques (éditions et rééditions des sermons de Massillon de 1745 à 1821), consignants des discours ayant été prononcés à haute voix et posant donc des problèmes spécifiques de « mise à l'écrit » (Régent-Susini 2017 ; Skupien Dekens 2019) ;

- ◆ un sous-ensemble, également graduel, des textes destinés à la consultation, non à une lecture publique, et donc adaptés à la **lecture silencieuse**. Il contiendra notamment (a) un noyau dur de **dictionnaires encyclopédiques** (*Tresor de recherches et antiquitez gauloises et françoises* de Borel (1655) ; *Dictionnaire de Broissinière* (1644 [1627]) ; *Grand Dictionnaire historique* de Moreri (1674) ; *Dictionnaire universel* de Furetière (1690) ; *Dictionnaire des arts et des sciences* de Th. Corneille (1694) ; *Dictionnaire de Trévoux* (1704) (b) élargi à d'autres d'écrits scientifiques, les **conférences** et **dissertations**, moins nettement destinées à un mode de consommation strictement visuel, car initialement lues et prononcées à haute voix – mais dans un cadre de communication formel, visant tout autant à garantir la réussite de l'acte de parole que la compréhension en lecture des énoncés ;
- ◆ un sous-ensemble, hérité et rémanent, d'**écrits multimodaux**, au rang desquels les pièces de théâtre de Racine et Molière, ayant fait l'objet d'une triple destination : la scène, où ces pièces étaient « déclamées » ; le hors-scène où, dans le cadre des salons littéraires, elles pouvaient être lues publiquement « à haute voix » ; le cabinet de travail où, comme le signale *a contrario* la préface de *L'Amour médecin* (« Tout le monde sait que les comédies sont faites pour être jouées, et je ne conseille de lire celle-ci qu'aux personnes qui ont des yeux pour découvrir dans la lecture tout le jeu du théâtre »), elles pouvaient faire l'objet d'une lecture silencieuse.

Parallèlement, STIH et CLESTHIA se proposent d'identifier et de réunir provisoirement, sous l'appellation de « **collections flottantes** », des ensembles de **textes non assignés, dépourvus de toute instruction de lecture explicite**, appartenant à des **genres de discours différents**, et dont les historiens du livre et de la lecture ont échoué à décrire, faute de témoignages, comment ils étaient lus (parmi eux : ensembles longitudinaux de **lettres privées** ; de **missives diplomatiques** ; de **gazettes** ; de **contes** – éditions et rééditions de ceux de Perrault notamment ; de **fables** – entre autres, éditions et rééditions de celles de La Fontaine.

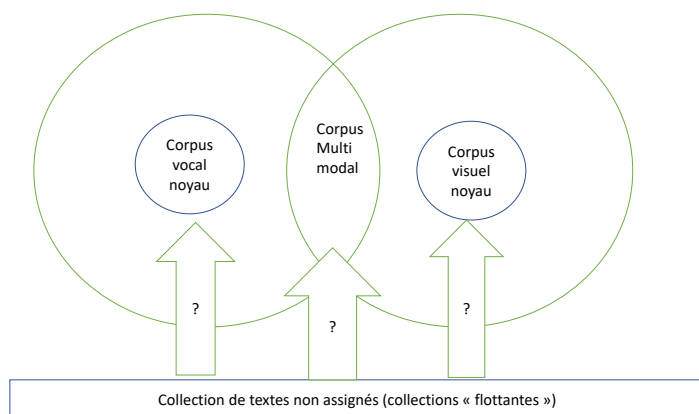


Fig. 2 : corpus et collections

Tout l'enjeu du présent projet est de parvenir à **documenter les lectures probables des éléments des collections « flottantes »**, par confrontation de leur architecture textuelle avec celle des textes des trois corpus principaux : « vocal », « visuel » et « multimodal ».

Pour pouvoir faire émerger ces « architectures » et les comparer, EcoLe recourra aux méthodes d'analyse, qualitatives et quantitatives, de la **linguistique de corpus**. Autant dire qu'il lui faudra **enrichir sous deux dimensions** (encodage des phénomènes textuels ; repérage des phénomènes typo-iconographiques) **ses corpus et les collections de textes non assignés**, en mobilisant les vertus combinées de l'annotation manuelle et du TAL, en fonction des observables (voir ci-dessous, I.c Objectif 2, section « Annotation » et section « Risques »).

### Objectif 3. – Analyse des corpus construits et des collections « flottantes »

Les enquêtes conduites dans ce cadre seront réunies dans un **ouvrage de synthèse**. Émanation d'une linguistique de corpus outillée, la *grammaire historique des pratiques de lecture* qui en sortira se veut :

- ◆ une *description intégrée* des opérateurs de mise en page, de segmentation du discours et de connexion, visant à déterminer si, suivant le mode d'appropriation (vocal, visuel, bimodal) auquel on les destinait, les textes du passé présentaient bien des architectures différenciées ;
- ◆ une *grammaire du changement*, mesurant, au sein de l'empan chronologique retenu (13<sup>e</sup>-18<sup>e</sup> siècles) :
  - ◆ l'incidence de la Révolution de l'imprimé – *ie*, du paramètre diamésique – sur le développement des pratiques de lecture et, partant, sur les formats textuels venus les porter. C'est cet objectif de « méso-diachronie » qui justifie (1) la présence de 2 équipes de diachroniciens, l'une spécialiste de la période de français médiéval et Renaissance, l'autre de français (pré)classique (2) et la mise en synergie de leurs résultats ;

- ◆ les foyers de résistance de la lecture bimodale/multimodale en régime (pré)classique, et de ses formats textuels afférents ;
- ◆ l'interaction des paramètres diachronique et diamésique avec ceux (1) des genres de discours – scientifique, littéraire, religieux, philosophique, juridique ; (2) des types de texte – narratif, descriptif, argumentatif, dialogal ; (3) des types formels – vers *vs* prose ; (4) des modes de diffusion éditoriale ; (5) des publics visés – public savant *vs* grand public.

L'objectif visé est d'aboutir à une modélisation souple (selon l'horizon de catégorisation graduelle développé dans la « théorie du prototype ») et dynamique (car soucieuse de mettre au jour des trajectoires diachroniques) du codage linguistique et typo-iconographique des pratiques de lecture ;

- ◆ une *histoire des idées linguistiques*. Une place sera faite aux imaginaires et représentations de l'unité « texte » en diachronie longue (13<sup>e</sup>-18<sup>e</sup> siècles). Cette étude sera confiée à la/ au doctorant.e financé.e par l'ANR. Elle mobilisera principalement un corps d'écrits métalinguistiques et métadiscursifs (grammaires, ouvrages de logique, recueils d'observations et de remarques, dictionnaires – généraux et spécialisés –, poétiques, traités de rhétorique, traités sur la langue, accessoirement, manuels typographiques...du domaine français), afin de pouvoir mesurer la part d'interaction, d'emprunt, de défi et d'hybridation réciproques entre les différentes traditions textuelles (tradition tabulaire *vs* linéaire). Les scènes de l'écriture dans les écrits littéraires – au sens large de « lettres humaines », englobant tous les domaines du savoir – seront également mobilisées. Chemin faisant, ces imaginaires et représentations du texte croiseront la question de la /des lecture(s). Cette documentation, neuve, afférente à des sources linguistiques et métalinguistiques peu regardées par les historiens du livre et de la lecture, devrait permettre d'enrichir la somme des témoignages externes aujourd'hui à disposition, et venir outiller, dans une perspective positiviste complétant l'option probabiliste ici retenue, le projet de grammaire historique des pratiques de lecture ;
- ◆ marginalement, une *grammaire contrastive*, documentant, dans le cadre d'un chapitre conclusif, via une boucle réflexive destinée à relier l'examen des formes du passé à celui des formes de notre présent, la multi-modalité des pratiques de lecture à époque ancienne, en la comparant à celle des pratiques françaises contemporaines. Le corpus émergent des **écrits inclusifs** et de leur double structuration (par le point médian, difficilement oralisable ; par la mobilisation de structures de coordination – « françaises, français » – ou de formes neutres – *iel* – vocalisables) constituera un observatoire pertinent, susceptible de faire apparaître les dynamiques visuelle *vs* vocale du /des français d'aujourd'hui.

## Objectif 4 – Expérimentations psycholinguistiques

Il s'agira de mettre à l'épreuve les hypothèses issues des travaux des diachroniciens et qui trouvent un écho dans les questions reliées à l'analyse psycholinguistique de la lecture, suggérant une adaptation de l'architecture textuelle en fonction de la lecture qui en sera faite. Pour cela, l'équipe SCALab évaluera si la modalité de lecture (à voix haute *vs* silencieuse) a un impact sur la compréhension que les lecteurs, apprentis et experts, ont du texte, et ceci en fonction de deux grands types d'architecture de texte, identifiés dans les opérations précédentes. L'hypothèse testée est que la compréhension sera meilleure lorsqu'il y a adéquation entre l'architecture textuelle et la modalité de lecture. En lien avec la littérature consacrée au domaine, il sera évalué, dans des conditions particulièrement contrôlées, si l'une des deux modalités de lecture, voix haute ou silencieuse, confère un avantage à la compréhension au sein des groupes considérés. Cette étude sera menée auprès d'enfants scolarisés en primaire, compte tenu de l'enjeu en termes de compréhension des mécanismes engagés dans l'apprentissage de la lecture. La question se présente sous un jour différent, mais tout autant fondamental, chez les lecteurs adultes, qui ont automatisé la lecture et sont familiers de la lecture silencieuse. Auprès de lecteurs dyslexiques, chez qui la question est centrale afin d'évaluer l'enjeu d'une prononciation (et donc ne permettant pas de contourner le simple traitement visuel des mots) dans la compréhension, de simples observations seront menées comme préalable à une étude plus complète proposée ultérieurement.

### I. B Positionnement par rapport à l'état de l'art

Situé à la croisée des **sciences de la culture**, des **sciences du langage** et de la **psycholinguistique**, ce projet est doublement novateur : de par **son ambition interdisciplinaire** tout d'abord, mettant en synergie, de manière inédite, linguistique diachronique et psycholinguistique ; ensuite, de par **les contributions** qu'il apporte à **chacun des trois domaines** mentionnés *supra*.

#### I. B.1 Positionnement du projet dans le champ des sciences de la culture

Depuis les années 1980, les historiens de la lecture (voir *supra*, I.A.1 et, pour une synthèse, Haug 2014), de la communication (Ong [1982] 2014 ; Vandendorpe 1999), du livre (Martin et Chartier 1982 et 1984 ; Martin 2000), de la culture (Zumthor 1987, Coleman 1996 ; Vitz 1999 ; Waquet 2003), ou encore, de la littérature (parmi d'autres, Dumonceaux 1990 ; Fumaroli et Goémé 1990 ; Noille 2004 ; Biet 2006 ; Bouchet 2008 ; Volpilhac 2015 ; Demonet 2016 ; Faugère 2020) se sont penchés sur le double régime de lecture qui prévalait en France et en Europe avant la Révolution française : celui, relativement nouveau, de la lecture privée, pour l'œil ; et celui, éprouvé, de la lecture publique, pour l'oreille.

Leurs analyses ont permis de penser l'articulation de la lecture « pour l'oreille » et de la lecture « pour l'œil » en termes (i) *diachroniques*, conférant à la période de l'Ancien Régime (16<sup>e</sup>-18<sup>e</sup> siècles) le statut de période transitoire, où le régime sonore dominant a basculé vers un régime silencieux (ii) *diamésiques*, avec le passage d'une culture manuscrite à une culture de l'imprimé (16<sup>e</sup> siècle) et du numérique (20<sup>e</sup> siècle), entraînant la « secondarisation » de l'oralité (Ong [1982] 2014) ; (iii) *diastématiques* et, partant, *situationnels* ou *variationnels* (voir *supra*), via l'identification des différents lieux de sociabilité (cours, salons, cénacles, places publiques,

scènes, églises) et des différents publics (savants *vs* peu lettrés ; adultes *vs* enfantins) où et à destination de qui s'exerçaient notamment « déclamation » et « haute voix ».

Pour autant, curieusement, dans la documentation de ces « scènes de lecture », si la question du texte, plus exactement, de son « architecture » (mise en page, ponctuation, mise en texte) a été travaillée (Ong, Chartier, Vandendorpe, Demonet), celle de sa corrélation avec les différentes pratiques de lecture n'a guère été abordée, du moins de manière systématique. La relation entre les architectures textuelles et les pratiques de lecture, l'incidence des premières sur la réussite des secondes, la part de risque dans le passage, pour un même texte, d'une pratique de lecture à une autre, n'ont pas été évaluées, si ce n'est par Baetens (2016), dont le travail, séminal, porte toutefois exclusivement sur la poésie moderne et contemporaine.

Du côté de la didactique, les très nombreux travaux sur la lecture ne sont pas non plus, pour la question intéressant EcoLe, d'un très grand secours. C'est qu'indépendamment de leur ancrage synchronique, après le virage technique des années 1970-1980, dont témoignent les « mesures de lisibilité » de Conquet et Richaudeau (1973) et les mises au point de Beaume (1987), ces travaux s'orientent plutôt, désormais, du côté du « sujet lecteur » (Vibert 2013), ou encore, de l'interaction entre le lecteur et le « contexte » de lecture (Potocki et Billotet 2020). La question du lien entre les caractéristiques formelles d'un texte et le succès *vs* insuccès de sa lecture par l'œil ou par l'oreille demeure très largement impensée. Et la redécouverte, depuis les années 1990, du bénéfice de la lecture à voix haute pour la mémorisation du lexique et de la syntaxe (Pled, Roudy et Hameau 1997 ; Ros-Dupont 1999) n'a pas modifié la donne : la lecture vocalisée est perçue dans la littérature comme un « après » de la lecture silencieuse, qu'elle suppose, et à laquelle elle vient se superposer. Elle en constitue, en quelque sorte, la version achevée. « Un bon lecteur à voix haute est aussi un bon lecteur silencieux » trouve-t-on dans les manuels scolaires dédiés à l'apprentissage de la lecture (*La compréhension en lecture. « Lire à haute voix au cycle 3 »*, Nathan pédagogie, 2020).

À tous ces titres, le présent projet, qui entend mettre à l'épreuve de la diachronie l'hypothèse selon laquelle certains écrits, de par leur structuration textuelle, se prêtent mieux que d'autres à telle ou telle pratique de lecture, est très largement disruptif.

### 1.B.2 Positionnement dans le champ des sciences du langage

Du point de vue de son ancrage linguistique, EcoLe s'inscrit dans une tradition de linguistique textuelle diachronique, que les membres des deux équipes STIH et CLESTHIA ont au demeurant contribué eux-mêmes à documenter, qu'il s'agisse (i) des travaux de Parussa (2020) – pour la période médiévale – et Maggi (2019) – pour la période moderne et contemporaine – sur l'autonomisation du code écrit par rapport au code oral, (ii) de ceux de Llamas Pombo (2020) ou encore, parmi d'autres, de Ducos (2019), Desbois-Tentile (2019), Halévy (2011) sur la ponctuation, (iii) de ceux de Badiou-Monferran (2020) sur les marqueurs de relation de discours, (iv) de ceux de Lombardero (2017) sur les chaînes de référence, (v) de ceux de Lefevre (2014) sur la phrase classique, (vi) de Badiou-Monferran (2019) et Skupien Dekens (2019) sur la segmentation du discours, (v) et, à la suite d'Arabyan (1994, 2012), de Lombardero (2018) sur les questions de mise en page et de Réach Ngô (2011) sur la dimension typo-iconographique de la textualité. L'apport majeur de EcoLe, à ce sujet, est de viser une description intégrée de l'ensemble des marqueurs de cohésion et de cohérence. Cette approche globalisante, rarement tentée en diachronie longue (même si les travaux fondateurs de Combettes sur le dynamisme communicationnel des états de langue anciens s'en approchent) constitue assurément l'un des points forts du projet. À terme, ce dernier devrait pouvoir proposer une modélisation dynamique des architectures textuelles, et constituer le pendant diachronique des travaux de linguistique textuelle conduits en synchronie, à une échelle globale, par Adam et/ou Charolles.

Plus ponctuellement, le chantier connexe, ouvert, de manière contrastive, sur les écrits inclusifs, invite dans le débat une problématique peu ou pas débattue par les linguistes synchroniciens, qui se sont emparé.e.s de cette question (parmi d'autres, Manesse et Siouffi 2019 ; Rabatel et Rosier 2019 ; Moirand 2021). Les travaux cités interrogent principalement la « dimension politique de l'écriture inclusive » (Moirand, à paraître : 4), ou ses fondements « historiques » (Siouffi et Manesse). Le mode de partage et de plaisir (par l'œil *vs* par l'ouïe) programmé par ses différents formats n'est pas (ou peu) documenté.

Enfin, pour revenir à la composante mère du programme, sa composante diachronique, l'un des apports majeurs du projet est sans doute qu'il propose des éditions semi-diplomatiques de ses corpus – par ailleurs très largement inédits (voir objectif 2). Si ce souci philologique s'inscrit pour le coup dans la droite ligne des principes et protocoles éditoriaux de la plate-forme *Sermo* (Skupien Dekens), il demeure somme toute assez rare dans les travaux de linguistique diachronique. À titre d'exemple, la *Grande grammaire historique du français*, ou encore le projet ANR DEMOCRAT, consacré aux chaînes de référence en diachronie longue, s'appuient sur des éditions modernisées. La convergence du scrupule philologique à l'œuvre dans les travaux d'édition numérique et de l'analyse linguistique modélisante constitue une puissante marque de fabrique, démarquant EcoLe de bien des projets auxquels il succède.

### 1.B.3 Positionnement dans le champ de la psycholinguistique

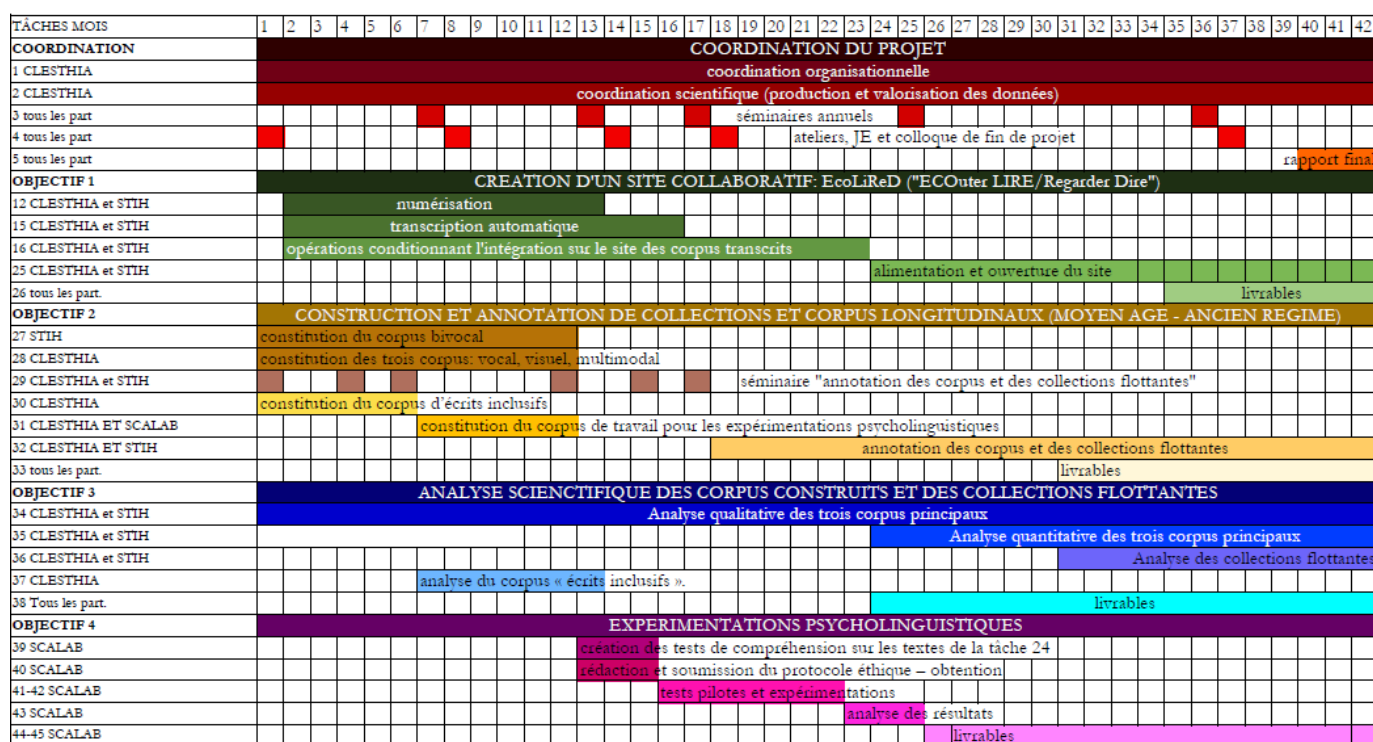
La recherche en psycholinguistique de la lecture s'appuie sur le point de vue consensuel selon lequel la compréhension en lecture résulte du produit de deux compétences fondamentales : l'identification des mots écrits, composante spécifique à la lecture, dont l'automatisation constitue un enjeu majeur des premières années d'apprentissage ; et la compréhension (Hoover & Gough 1990). Dans cette équation, la compréhension écrite (le produit) n'est pas supposée différer fondamentalement de la compréhension orale (l'un des deux composants) au plan des mécanismes de traitement engagés, mais elle se trouve modulée par la précision et la rapidité mises en œuvre dans la reconnaissance de mots (Massonnié *et al.* 2019). Ce « modèle simple » (*Simple View*) a servi de base aux recherches, même si on admet que la situation est probablement plus complexe (Ouellette & Beers 2010). Afin d'éprouver le poids de l'automatisation de la reconnaissance de mots dans la compréhension en lecture, plusieurs études ont comparé la compréhension de phrases entendues *versus* lues (Casalis *et al.* 2013). On peut retenir que c'est à partir de la 5<sup>e</sup> année de primaire que compréhension orale et compréhension écrite deviennent équivalentes (Lecocq *et al.* 1996). Ces travaux permettent également d'évaluer l'impact des indices morphosyntaxiques muets sur la compréhension à l'écrit (e.g. certaines marques du pluriel visibles et non audibles peuvent

faciliter l'attribution des rôles syntaxiques). Soulignons que dans ces études, la compréhension de l'écrit est évaluée après une lecture silencieuse (e.g. Berninger & Abbott, 2010). Dès lors que la reconnaissance des mots est automatisée, on considère souvent que la lecture à voix haute ralentit le rythme de lecture, ce qui pourrait altérer la compréhension (Kriebler *et al.* 2017). Dans ces conditions, mais pour des raisons finalement assez peu justifiées, les études ne considèrent la compréhension en lecture qu'en modalité silencieuse. Parallèlement, les travaux sur l'identification de mots ont mis en évidence l'importance décisive du codage phonologique appliqué au mot écrit. En effet, compte tenu de ce qu'est le système alphabétique, la mise en relation des mots écrits avec leur forme orale passe par une analyse fine des processus de codage phonologique. Ce processus de codage phonologique est essentiel dans la constitution d'un lexique orthographique écrit (Share, 1995). Cela dit, alors qu'on a pu montrer qu'un code phonologique est automatiquement activé en lecture silencieuse (à l'insu du lecteur, Sauval *et al.* 2017), les liens qu'entretiennent ce code phonologique et l'activation phonologique ouverte en lecture à voix haute ne sont pas encore complètement connus.

Quelques travaux, peu nombreux, ont cherché à savoir si la compréhension d'un texte différait selon qu'il était lu silencieusement ou à voix haute. Les études ne permettent pas de trancher en faveur d'une modalité : si la plupart des études ne trouvent pas de différences (Garcia-Rodicio *et al.* 2018 ; Hale *et al.* 2011), certaines en revanche mettent en évidence des scores de compréhension supérieurs en lecture à voix haute (Hale *et al.* 2007). Notons que certaines études suggèrent que les processus engagés pourraient différer (O'Brien *et al.* 2014, Vorstius *et al.* 2014), avec une situation de « double-tâche » en oralisation, potentiellement pénalisante. Cependant, ces études n'ont pas permis d'évaluer le lien qui pourrait exister entre une architecture textuelle et la modalité de lecture : en général, les mêmes textes, ou des textes tout à fait comparables étaient utilisés dans les deux modalités de lecture. On peut penser *a priori* que ces textes étaient représentatifs d'une architecture dévolue à la lecture silencieuse. Les études menées dans le cadre du présent projet pourraient ainsi, si une tendance nette se dégageait, permettre de conclure à l'incidence de la modalité de lecture sur la compréhension. Mais la question de l'adéquation de la modalité de lecture à l'architecture fonctionnelle reste totalement ouverte. À dire vrai, cette question n'est pas considérée dans la littérature psycholinguistique à notre connaissance, et c'est la collaboration avec des diachroniciens de la langue, sensibles à l'évolution de la forme des textes, qui a permis l'émergence de cette question. En effet, la question du rôle de l'architecture formelle du texte a bien été évaluée en psycholinguistique, mais toujours en lien avec la compréhension et la capacité à identifier les informations (Goldman & Rakerstraw 2000), jamais en lien avec la modalité de lecture (silencieuse ou voix haute) qui est réalisée, puisque la lecture est pratiquement toujours réalisée silencieusement (sauf chez les tous-débutants en lecture).

## I.C Méthodologie et gestion des risques

Le programme de travail est figuré dans le diagramme de Gantt ci-dessous :



**Coordination.** Le nombre de collaborateurs important au sein du projet (28), son consortium trans-séculaire et interdisciplinaire (voir ci-dessous, II.a), l'interaction continuée de sa composante « humanités numériques » (objectif 1) avec les autres composantes du projet – composante TAL (objectif 2), composante expérimentale (objectif 4), composante linguistique et philologique (objectifs 2 et 3) – et le triple champ d'application de ses opérations de valorisation (auprès des communautés scientifiques, des enseignants des premier et second degrés, du grand public) supposent une coordination organisationnelle et scientifique attentive (tâches ① et ②,



sous la responsabilité de Claire Badiou-Monferran). **Deux outils** seront mis en place pour aider à fédérer la communauté de travail : [1] une bibliographie en accès réservé via **Zotero** ; [2] un **Sharedocs** en accès ouvert, constituant le site de dépôt et d'échange de toutes les documentations nécessaires à la recherche, à la pédagogie et à la valorisation. **Un séminaire** d'une à deux journées annuelles, coordonné par Claire Badiou-Monferran, Séverine Casalis et Gabriella Parussa, sera ouvert à l'ensemble des participants. Il procédera à la mise à jour de l'avancée des travaux et des difficultés rencontrées, et engagera une réflexion épistémologique sur le projet, ses réflexions et ses choix (④, mois 7, 13, 17, 25, 36). Il sera accompagné de deux **ateliers de travail** et de deux **journées d'études** en présentiel, visant la constitution d'une **culture commune** et d'un **langage commun** – Atelier<sub>1</sub> : « Kick off », organisé par CLESTHIA (⑤, mois 1) ; Atelier<sub>2</sub> à la Sorbonne nouvelle : état de la question sur « Les arts de lire en France et en Europe avant la Révolution » (⑥, mois 8) ; JE<sub>1</sub> organisées par SCALab à Lille : « Pluridisciplinarité. Interdisciplinarité, Transdisciplinarité : vers une psycholinguistique diachronique ? » (⑦, mois 14) ; JE<sub>2</sub> organisées par STIH à Paris : « Annoter les corpus anciens » (⑧, mois 18). Coordonné par les trois équipes, un **colloque de fin de projet** portant sur « Les mutations du texte et de la lecture en France et en Europe avant la Révolution » mettra en synergie les acquis trans-séculaires et interdisciplinaires du programme de recherche EcoLe (⑨, mois 37). Le rapport de fin d'ANR (tâche ③, mois 40-42) sera élaboré collectivement (contribution de tous les scientifiques).

## Objectif 1 – Création du site collaboratif : **ÉcoLiReD** (« ÉCOuter LIRE/REgarder DIRE »)

**Responsables** : [1] Anne Réach-Ngô (ARN), référente « humanités numériques » (ITEM, rattachée pour le projet à l'équipe CLESTHIA) ; [2] Gaël Lejeune (GL), référent TAL – traitement automatique des langues (STIH) ; [3] Richard Walter (RW), Ingénieur de recherche, CNRS (UMR THALIM, Sorbonne Nouvelle), coordonnateur de la plate-forme EMAN.

RW et ARN encadreront les deux ingénieur.e.s d'études (IE) recruté.e.s respectivement, sur financement de l'ANR, pour 12 mois, l'un.e en début de projet (mois 2-13), l'autre en fin de projet (mois 24-35).

**Partenaires** : tous les scientifiques de CLESTHIA (Sorbonne Nouvelle) et STIH (Sorbonne Université) ; + Séverine Casalis, responsable de l'équipe SCALab (Lille).

**Méthodologie / Détail des tâches.** Les tâches de cet objectif peuvent être appréhendées sous 4 dimensions : 1. numérisation ; 2. transcription automatique ; 3. construction et 4. alimentation du site. Afin d'éviter tout goulot d'étranglement, les tâches afférentes à 1, 2 et 3 seront réalisées en parallèle.

### 1. Numérisation [GL – tous les scientifiques concernés de CLESTHIA et STIH]

**Mois 2 à 13** – La numérisation des manuscrits et imprimés à verser dans les corpus d'EcoLe sera confiée à un prestataire ⑫<sub>1</sub>. Le contrôle qualité sera assuré, concomitamment, par l'ensemble des scientifiques CLESTHIA et STIH ⑫<sub>2</sub>.

### 2. Transcription automatique [GL, ARN, RW, IE – avec la coll. des membres de CLESTHIA et STIH concernés]

**Mois 2-17** – Le processus de transcription automatique se décomposera comme suit : – État de l'art effectué par RW, GL et le/la premier.e IE (CDD de 12. Mois 2-13) sur les différents logiciels de reconnaissance de caractères (HTR et OCR) à disposition ⑬ – Parallèlement, sélection par les scientifiques de CLESTHIA et STIH d'un corpus d'entraînement sur la base de numérisations d'ores et déjà existantes, à verser dans les corpus du projet ⑬ – Phase d'entraînement/ré-entraînement des outils d'HTR/OCR retenus, sous l'égide GL et RW ⑭ – Transcription automatique des témoins manuscrits et imprimés du corpus. Amélioration du moteur de reconnaissance en continu. Relecture et de validation des résultats par tous les scientifiques concernés de CLESTHIA et STIH, sous la coordination d'ARN et GL ⑮.

### 3. Opérations conditionnant l'intégration sur le site des corpus transcrits [les mêmes, avec la collaboration de C. Badiou-Monferran, A. Gerstenberg, G. Parussa et C. Skupien Dekens]

**Mois 1-18** – Mise en place de la chaîne de traitement des données. R. Walter et A. Réach Ngô ⑩. – Établissement de la méthodologie d'encodage : préparation du schéma de données pour le traitement en XML-TEI, notamment pour l'étiquetage morphosyntaxique et l'enrichissement sémantique ⑯ – Préparation de l'environnement de travail : développement des fonctionnalités – plug-in – pour l'intégration des outils textométriques sur EMAN, avec phase de test. Les mêmes ⑰ – Avec C. Badiou-Monferran et G. Parussa, création des métadonnées (métadonnées Dublin core et IIIF) ⑱ – Intégration des données venant des tâches précédentes – numérisation ; reconnaissance ; vérification ⑲. Avec le concours de C. Skupien Dekens, responsable du projet SERMO (Neuchâtel), et A. Gerstenberg, responsable du projet APWCF (Berlin/Potsdam), étude des possibilités d'interopérabilité d'EcoLiReD avec les corpus et les outils de traitements produits par les sites partenaires ⑳ – Mise en place de place de l'interopérabilité avec les corpus et les outils de traitements des sites partenaires ㉑ – Personnalisation du site EcoLiReD en fonction des résultats des autres tâches ㉒.

### 4. Alimentation du site — Ouverture du site [RW, IE – avec la collaboration des scientifiques de CLESTHIA, STIH et S. Casalis pour SCALab]

**Mois 23-42** – Recrutement du la/le nouvel.le IE (CDD de 12 mois, mois 24-35), par R. Walter, C. Badiou et G. Parussa ㉓ – Intégration des contenus produits dans le projet afin de permettre leur annotation par l'équipe. Alimentation du site EcoLiReD en continu par la /le IE. L'intégration se fera au fil de l'eau, tout au long des 24<sup>e</sup>-42<sup>e</sup> mois. Pour cette opération, R. Walter prendra le

relais de l'IE parvenu.e en fin de contrat à partir du 36<sup>e</sup> mois ②4. – À partir de la deuxième année, ouverture du site (accessible auparavant par accès réservé aux collaborateurs du projet) aux communautés scientifiques (mois 24-29). Ajustement de l'interface d'interrogation et de publication après les premiers retours d'utilisateurs ②51. – À partir de la troisième année, ouverture du site aux enseignants, aux scolaires, au grand public (mois 35-42). Prise en compte au fil de l'eau des retours d'utilisateurs ②52. – Avec C. Badiou-Monferran, S. Casalis et G. Parussa, mise à disposition des livrables destinés à l'ensemble de ces publics (mois 35 à 42) ②6.

### Livrables

- ◆ EcoLiReD fournira un très vaste corpus (a) au format XML TEI, constitué de trois sous-corpus, intégralement transcrits selon un protocole semi-diplomatique, assortis de métadonnées, annotés et interrogeables (b) et restituant les caractéristiques typographiques des documents originaux (conformément à une utilisation originale d'Omeka, en tant que bibliothèque numérique plaçant au cœur de la démarche d'éditorialisation le document image).
- ◆ Il intégrera tout ou partie des corpus déjà constitués dans le cadre d'autres projets, déposés sur le Réseau (RCFC), mais qu'il enrichira : (a) en proposant une mise à jour semi-diplomatique de la transcription moderne actuellement disponible des deux corpus « Bailly » et « APWCF » (avec reprise de la ponctuation et de la mise en page originelles) ; (b) en augmentant le nombre de textes (pour le corpus du site SERMO, faisant déjà l'objet d'une transcription semi-diplomatique).
- ◆ EcoLiReD (a) mettra à disposition des livrets de haute vulgarisation (conçus par CLESTHIA, SCALab et STIH) pour les enseignants ; (b) proposera des jeux interactifs conçus par SCALab à destination des scolaires et du grand public ; (c) hébergera une exposition virtuelle (livrable de CLESTHIA et STIH) [textes anciens et iconographie] ; (d) livrera des enregistrements audio de lectures à haute voix (CLESTHIA, SCALab, STIH).

### Risques et gestion des risques

Les risques identifiés pour cet objectif 1 sont les goulots d'étranglement. Ils seront jugulés par la parallélisation des tâches.

#### Plan de gestion des données – Coordination : Anne Réach-Ngô. APPLICABLES AUX OBJECTIFS 1, 2, 3, 4

Pour mener à bien le projet, un plan de gestion des données (PGD) sera fait avec l'outil Opidor et permettra de gérer le flux des données entre les différentes étapes et lieux du projet. Pour la réalisation du site ECOLIRED, un plan de gestion de données plus spécifique sera réalisé en se basant sur celui plus général de la plateforme EMAN. La question d'un entrepôt sera également étudiée (en particulier avec Zenodo, souvent utilisé en psycholinguistique).

## Objectif 2 : construction et annotation de nouveaux corpus

**Responsables :** Claire Badiou-Monferran (Sorbonne Nouvelle) et Carine Skupien Dekens (Neuchâtel, rattachée à CLESTHIA dans le cadre d'EcoLe) pour CLESTHIA ; Gabriella Parussa (Sorbonne Université, référente STIH) et Gaël Lejeune (STIH) pour STIH.

Carine Skupien Dekens, responsable du projet FNS SERMO, achevé en 2018 – analyse et exploitation semi-automatique d'un corpus longitudinal (1550-1750) de sermons protestants, comprenant en l'état 600571 tokens (<http://sermo.unine.ch/SERMO/>) – co-animera le séminaire de travail EcoLe (mois 1 à 17) en vue du partage des résultats et des outils de SERMO (chaîne de traitement et des scripts ; lexiques enrichis ; annotation de corpus ; interface internet).

Gaël Lejeune, taliste, encadrera la /le post-doc en charge l'annotation syntaxico-sémantique et typo-iconographique des corpus. Cette dernière, rattachée à STIH, sera recrutée pour 24 mois à partir du mois 18 sur la base d'un financement ANR.

Claire Badiou-Monferran et Gabriella Parussa coordonneront l'ensemble des opérations scientifiques réalisées par les membres de CLESTHIA et STIH.

**Partenaires :** membres de CLESTHIA, de STIH – et en plus, pour les tâches ③1 et ③3, membres de SCALab.

**Méthodologie/ Détail et déroulé des tâches.** EcoLe se propose d'inventer de nouveaux types de corpus ; de les construire ; de les annoter pour permettre d'en produire des analyses linguistiques ; et d'en faire paraître une édition numérique semi-diplomatique, étiquetée et interrogeable sur EcoLiReD.

### 1. Constitution des trois corpus principaux [STIH et CLESTHIA]

**Mois 1-12** – En se fondant sur la **collecte d'instructions de lecture explicites** (« liront ou écouterons » vs « prononceront à haute voix » vs « liront silencieusement ») alimentée, entre autres, par les enquêtes du/de la doctorant.e financée par l'ANR, l'équipe STIH, d'un côté, consacra la première année de l'ANR à la stabilisation des éléments (témoins manuscrits et/ou imprimés) à verser à son **corpus bimodal** ②7 ; l'équipe CLESTHIA, de l'autre, construira ses **deux corpus différenciés**, « visuel » vs « vocal ». Parallèlement, CLESTHIA identifiera, en diachronie, du point de vue, principalement, **des genres de discours** et toujours à partir du jeu d'étiquettes disponible, les **foyers de résistance de la lecture multimodale**. Cet ensemble viendra augmenter le corpus bimodal de l'équipe STIH, dans une perspective longitudinale ②8. Dans le même temps, en vue de la tâche 37 ci-dessous, l'équipe CLESTHIA réunira des collections de textes appartenant à des genres de discours différents (lettres privées – responsable : Myriam Bergeron-Maguire – ; missives diplomatiques – collection/corpus APWCF d'Annette Gerstenberg et collection Bailly d'Antonella Amatuzzi – ; fables et contes – responsable : Claire Badiou-Monferran) dont le mode de lecture (bimodal ? oralisé ? silencieux ?), non attesté, est encore

aujourd'hui réputé inassignable. Hétéroclite par définition, cet ensemble, se distinguant d'un « corpus », constituera tout au mieux un agrégat de **collections de textes non assignés**, en attente de traitement (**collections « flottantes »**) <sup>28</sup><sub>bis</sub>.

## 2. Constitution des 2 corpus de travail : corpus de contraste [CLESTHIA] – corpus de validation [CLESTHIA et SCALab]

**Mois 1-6** – Construction du **corpus de contraste**, permettant de comparer la multi-modalité des pratiques de lecture à époque ancienne avec celle des pratiques françaises contemporaines. Collecte d'écrits inclusifs allant du plus codifié au moins codifié : (i) sélection d'écrits institutionnels émanant des signataires de la « Convention d'engagement pour une communication publique sans stéréotype de sexe » (<https://www.haut-conseil-egalite.gouv.fr/stereotypes-et-roles-sociaux/actualites/article/liste-des-signataires-de-la-convention-d-engagement-pour-une-communication>) (ii) courriels d'universitaires en situation de communication (a) formelle *vs* (b) informelle (iii) fictions littéraires (Alpheratz, *Requiem*, 2015 ; textes déposés sur le site « Ateliers de l'antémonde », notamment, *Frictions* : <https://antemonde.org/textes/frictions/>) La tâche sera confiée à E. Lombardero (CLESTHIA), jeune chercheuse ayant commencé à développer un champ d'expertise spécifique autour de l'histoire du féminin dans la langue <sup>30</sup>.

**Mois 7-12** – Construction du **corpus de validation : corpus de travail des psycholinguistes**. Il sera le fruit d'une collaboration entre CLESTHIA et SCALab, et fournira à l'équipe des psycholinguistes un groupement de textes lui permettant d'avoir accès aux trois formats d'architecture textuelle – notamment les formats vocal et bimodal, spécifiques (tout du moins, caractéristiques) selon notre hypothèse des écrits du passé. En l'état, à cause de la trop grande différence linguistique, morphosyntaxique notamment, entre le français médiéval et le français d'aujourd'hui, l'équipe STIH ne participera pas à cette tâche. La collaboration de CLESTHIA et SCALAB aboutira à la sélection, validée par les deux équipes, de deux ensembles de textes de français (pré)classique, prenant en compte tous les formats textuels possibles (visuel, vocal, bimodal) :

- ♦ 1<sup>er</sup> ensemble : ensemble de **12 textes de français préclassique/ou classique, d'environ 1 page, explicitement destinés, pour 6 d'entre eux, à une lecture à haute voix, et pour les 6 autres, à une lecture silencieuse**. Le nombre 12 permettra de faire tourner les 4 conditions de lecture (a. lecture à haute voix d'un texte programmé pour une lecture à voix haute ; b. lecture à haute voix d'un texte programmé pour une lecture silencieuse ; c. lecture silencieuse d'un texte programmé pour une lecture silencieuse ; d. lecture silencieuse d'un texte programmé pour une lecture à haute voix) en renouvelant 3 fois l'expérience – afin de neutraliser les bruits potentiels dus aux spécificités/difficultés d'un texte donné.
- ♦ 2<sup>e</sup> ensemble : ensemble de **16 textes bimodaux de français préclassique/ou classique, d'1 page environ, présentés, via des adaptations de mise en page et de ponctuation principalement, dans deux formats** : l'un plutôt **visuel**, l'autre plutôt **vocal**. Le nombre de 16 permettra de faire tourner les deux formats de présentation d'un même texte dans chacune des deux situations de lecture (format visuel *vs* format vocal en situation de lecture silencieuse ; format vocal *vs* format visuel en situation de lecture à haute voix) et de reconduire l'expérience quatre fois, avec des textes différents, afin, là encore, de neutraliser les bruits potentiels dus aux spécificités/difficultés d'un texte particulier

La **sélection des textes** et leur **réécriture/adaptation** destinée à lever les difficultés de compréhension liées au recul temporel (modernisation de l'orthographe, chasse aux faux-amis et au lexique archaïque, adaptation morphosyntaxique locale, etc., dans le respect de la mise en page et de la ponctuation d'origine) sera effectuée, sous l'égide de C. Badiou-Monferran, par les membres de CLESTHIA rompus à l'exercice de la transposition, parce que traductologue (L. Maggi), traductrice (E. Lombardero) et/ou éditrice de textes anciens dans des collections Grand public (A. Régent Susini) <sup>31</sup>.

## 3. Annotation des corpus

**Les trois corpus principaux** seront enrichis sous deux angles : **encodage des phénomènes textuels** (ponctuation, connecteurs, isotopies, chaînes de référence...) et **repérage des phénomènes typo-iconographiques** (mise en espace des textes, identification des ornements, hiérarchisation des titres, *marginalia*, alternance des fontes de caractères...) concourant à la structuration des écrits. Du 3<sup>e</sup> au 17<sup>e</sup> mois, C. Skupien Dekens (voir ci-dessus) et F. Lefeuvre, CLESTHIA (tout à la fois spécialiste du 17<sup>e</sup> siècle et du 21<sup>e</sup> siècle et qui, à partir de son corpus CFPP<sub>2000</sub>, a développé des compétences particulières en matière d'annotation), modèreront, avec l'appui de Gaël Lejeune, une série de visioconférences et/ou d'ateliers de travail, afférents à l'identification et à la construction de l'étiquetage visé <sup>29</sup>. Pendant les **mois 18 à 42**, Gaël Lejeune encadrera la/le post-doc, recruté.e à STIH sur financement de l'ANR. Ce.te dernier.e sera en charge de l'annotation syntaxique, sémantique, énonciative, typo-dispositionnelle et iconographique des trois corpus principaux, ainsi que de celle des collections de textes non assignés. Elle/il sera aidé.e dans cette tâche par la/le doctorant.e contractuel.le, un.e vacataire, et tous les scientifiques de CLESTHIA et STIH concernés <sup>32</sup>. L'ensemble des travaux ci-dessus mentionnés donnera lieu à différents livrables <sup>33</sup>.

### Livrables

- ♦ **Édition numérique des collections de textes non assignés**, également annotés – venant compléter l'édition numérique des trois corpus annotés (voir ci-dessus, **objectif 1, Livrable**). Ces dernières seront déposées sur EcoLiReD, non comme « corpus », mais comme « cargaisons flottantes », en attente de destination. L'annotation des documents les constituant (Lettres privées, Missives diplomatiques, Fables, Contes) devrait permettre aux utilisateurs du site (les collaborateurs du projet, mais aussi, à termes, les chercheurs intéressés) de faire apparaître, pour chaque élément de l'ensemble, un principe de structuration spécifique – visuel, vocal, bimodal – et donc, de documenter pour chacun des textes, par comparaison avec l'architecture textuelle des corpus principaux, un mode de lecture probable. Les collections seront déposées telles quelles sur le site, pour en permettre l'interrogation et l'exploitation dans le cadre d'autres projets.

- ◆ **Édition critique de certains des documents versés à l'un des trois corpus principaux** : Pierre Borel, *Tresor de recherches et antiquitez gauloises et françoises reduites en ordre alphabetique*, Paris, Augustin Courbé 1655, éd. A. Amatuzzi ; Antoine Furetière. *Dictionnaire universel* éditions méconnues de 1692 et 1694, éd. C. Badiou-Monferran. [Mise à disposition de témoins méconnus et/ou d'éditions critiques manquantes].
- ◆ **Édition numérique** sur EcoLiReD, dans le respect des principes FAIR, du **corpus de travail des psycholinguistes**, élaboré avec les membres de CLESTHIA, mais dans le cadre d'un dispositif « en miroir », mettant en regard le document original, édité diplomatiquement, et son adaptation pour les expérimentations.
- ◆ **Production d'écrits scientifiques**, en français et en anglais, **dans des revues internationales du meilleur niveau** sur (a) le corpus d'écrits inclusifs, (b) le travail collaboratif et interdisciplinaire des diachroniciens et des psycholinguistes, (c) les difficultés rencontrées dans l'annotation syntaxico-sémantique et typo-iconographique des corpus anciens. Une attention particulière sera portée à **l'accès ouvert immédiat de ces publications (voie dorée, diamant ou verte en licence CC-BY)**.

## Risques

Les risques identifiés pour cet objectif tiennent principalement aux **limites actuelles de l'annotation syntaxico-sémantique et typo-iconographique en TAL**. En effet, certains observables (ponctuation, connecteurs, saut de ligne, alinéas, titulatures...) sont bien documentés (voir entre autres les modèles MAT – Modèles d'Architecture Textuelle – de Virbel 1989, Pascual 1996, Maurel *et al.* 2003) et leur identification bien outillée (voir par exemple les initiatives Inex sur la structuration des livres, Doucet *et al.* 2013, ou plus récemment la campagne d'évaluation FinToc sur la structuration des documents financiers). Mais pour d'autres observables comme l'identification des « phrases » et des « périodes » ou encore la caractérisation précise des styles et informations visuelles (graisse, italique, colonnage ...), l'automatisation demeure plus incertaine notamment parce que les outils de TAL existants ont tendance à écraser les observables pour se concentrer sur des tâches très précises. À ce titre, le projet EcoLe, visant un traitement global et complexe des architectures textuelles, invite le TAL (et l'équipe CERES de Gaël Lejeune en particulier) à un changement d'échelle, et constitue un vrai défi pour la recherche des cinq prochaines années. Dans ce cadre, les risques, qui ne remettent pas en question la faisabilité globale de l'annotation requise, cette dernière reposant sur un grand nombre d'acquis, seront localement jugulés par la possibilité du recours à une **annotation manuelle** (post-doc de 24 mois attiré.e) pour les situations où la valeur ajoutée offerte par l'outil informatique est plus incertaine.

## Objectif 3 – Analyse des corpus construits et des collections « flottantes »

**Responsables** : Claire Badiou-Monferran (CLESTHIA) et Gabriella Parussa (STIH)

**Partenaires** : tous les scientifiques de CLESTHIA et STIH

### Méthodologie / Détail des tâches

**Pour l'analyse des trois corpus principaux**, dont le mode d'appropriation (bimodal, vocal, visuel) est documenté, la démarche adoptée relèvera d'une approche **Corpus based** (Biber 2015) : approche qui utilise les corpus pour confirmer ou infirmer une hypothèse, en l'occurrence, l'hypothèse selon laquelle les textes destinés à la haute voix et ceux destinés à la lecture silencieuse ne mobilisent pas (ou pas exactement, dans le cas des documents bimodaux) la même architecture textuelle. Ces corpus feront l'objet d'**analyses qualitatives** (mois 1 à 42 : 34) et **quantitatives**, une fois annotés et déposés sur le site (mois 24-42 : 35). Inversement, **l'analyse des « collections flottantes »** – collections de textes non assignés, dont on ignore comment ils étaient lus – s'appuiera pour sa part sur une démarche inductive : **démarche Corpus driven** (Biber 2015), qui exploite les données des corpus sans *a priori*. Elle adoptera une **approche comparative** : déduction et reconstruction des modes de lecture probables par confrontation de l'architecture textuelle de chacun de ses éléments avec celle des textes des trois corpus principaux, « vocal », « visuel » et « multimodal » 36. Dans le cadre de cette **linguistique de corpus**, et pour toutes les tâches 34-36, afin de proposer une modélisation fine des différentes architectures textuelles, il sera tenu compte de des **paramètres variationnels** suivants : 1. variation temporelle ; 2. variation situationnelle (2.1 public visé : savant *vs* peu lettré ; 2.2 cadre de communication : public *vs* privé ; collectif *vs* individuel ; 2.3 nature de la communication : formelle *vs* informelle) ; 3. variation générique (genres de discours : scientifique, littéraire, religieux, philosophique, juridique) ; 4. variation textuelle (types de texte : narratif, descriptif, argumentatif, dialogal) ; 5. variation formelle (prose *vs* vers) ; 6. Variation diamésique (6.1 supports : manuscrits *vs* imprimés – ; 6.2 format ; 6.3 modes de diffusion). L'analyse du corpus contemporain d'écrits inclusifs pourra être produite dès sa constitution, ie, à partir du 7<sup>e</sup> mois 37. Sont attendus les livrables ci-dessous 38.

### Livrables

- ◆ Un **ouvrage collectif** à paraître en édition papier dans une maison d'édition à comité de lecture international : *S'écouter lire. Vers une grammaire historique des pratiques de lecture (Domaine français. 13<sup>e</sup>-18<sup>e</sup> siècles)*, incluant un chapitre liminaire ou conclusif sur les pratiques de lecture contemporaines, à partir de l'observatoire des écrits inclusifs.
- ◆ Un **mémoire de thèse** sur les « Imaginaires, représentations et pratiques du texte en France avant la Révolution française ».
- ◆ Des **publications** en anglais et en français dans des revues internationales du meilleur niveau, présentant et développant les analyses de l'ouvrage collectif de synthèse *supra*. (**Open accès**, voies dorée, diamant ou verte avec *postprints* déposés immédiatement sur HAL)
- ◆ Des instruments de **valorisation** : **communications, conférences** nationales et internationales, notamment dans les **pays partenaires** (Allemagne, Espagne, Italie, Suisse).

**Risques** : aucun risque intrinsèque.

## Objectif 4 – Expérimentations psycholinguistiques

**Responsable** : Séverine Casalis (SCALab)

**Membres** : équipe SCALab. Ludivine Javourey-Drevet ; Gwendoline Mahé ; Laurent Sparrow.

Séverine Casalis encadrera le/la post-doc recruté.e via un financement de l'ANR pour une durée de 18 mois (Mois 13-29)

**Méthodologie et détail des tâches.** **Rappel de l'objectif** – Il s'agira de tester l'hypothèse selon laquelle l'impact de la modalité de lecture (à voix haute *vs* silencieuse) sur la compréhension du texte lu dépend de l'architecture du texte (architecture visuelle, de type « tabulaire » *vs* architecture vocale, de type « linéaire »). Plus précisément, on testera le lien entre le degré de compréhension en lecture, et l'adéquation entre le format de présentation (l'architecture textuelle) et la modalité de lecture. Seront évaluées : (i) la vitesse de lecture (ii) la compréhension (questions littérales et inférences connectées au texte et basées sur les connaissances). **Participants** – Deux types de populations participeront à l'expérience : (a) des **apprentis lecteurs** ayant franchi les premières étapes du décodage en lecture (CM2). Pour eux, on fera l'hypothèse d'une meilleure compréhension en lecture lorsqu'il y a adéquation entre la modalité de lecture et le format de présentation ; (b) des **lecteurs experts (étudiants)**. À leur sujet, l'hypothèse d'un meilleur niveau de compréhension lorsqu'il y a adéquation entre le format textuel et la modalité de lecture sera complétée par une hypothèse, liée à une perte de familiarité avec la lecture à voix haute, de supériorité de la compréhension dans la lecture silencieuse en règle générale. Le premier groupe sera composé d'une cinquantaine d'enfants (INSPE, Rectorat, IEN, DASEN) dont on aura vérifié qu'ils sont à l'heure scolairement ; qu'ils présentent un niveau de raisonnement cognitif non verbal normal (test de Raven, 1938) ; qu'ils présentent un niveau de lecture (décodage) normal (test de l'Alouette, Lefavrais [1967] 2005) ; et qu'ils présentent un niveau de compréhension normal (ECOSSE, Lecocq, 1996). Le deuxième groupe comprendra une centaine d'étudiants en psychologie (50 pour la première étude, 50 pour la seconde). Ainsi qu'indiqué dans l'objectif, nous mènerons également une observation pilote auprès de quelques enfants dyslexiques consultant en cabinet d'orthophonie, mais ceci ne rentre pas dans ce cadre expérimental rigoureux et sert d'observations préalables. **Méthode** – Les deux approches présentées *supra* (objectif 2, tâche ③①) seront mobilisées : la première, consistant à faire tourner les quatre modalités de lecture, sera appliquée aux deux populations de lecteurs ; la seconde, consistant à faire tourner les différentes architectures textuelles des textes bimodaux selon les deux modalités de lecture (silencieuse *vs* voix haute) fera l'objet d'une expérience spécifique, supplémentaire, appliquée au seul deuxième groupe de lecteurs experts, pour des raisons de moyens. Chaque participant lira 3 textes (4 pour les participants de la seconde étude chez l'expert) dans chacune des 4 conditions (12 et 16 respectivement au total) ; l'affectation des textes à la modalité de lecture (voix haute ou silencieuse) sera contrebalancée. Le plan choisi sera donc en intra-sujet. **Évaluations** – L'évaluation du degré de compréhension en lecture crociera plusieurs critères : (i) temps de lecture total (ii) nombre d'erreurs de lecture en situation de lecture à voix haute (iii) qualité des réponses à un questionnaire contenant 10 questions (4 littérales ; 3 relatives aux inférences textuelles, « *text connecting inferences* » – anaphores et relations de discours ; 3 relatives aux inférences basées sur les connaissances). Chaque condition fera l'objet d'une notation conduisant de 1 à 30 (40 pour la 2<sup>e</sup> expérience chez l'expert). **Déroulement chronologique des tâches** – Après la constitution du corpus de travail avec les membres de CLESTHIA (voir *supra*, objectif 2 ③①, mois 7-12), Séverine Casalis et Ludivine Javourey-Drevet procéderont à l'élaboration des tests de compréhension (tâche ③②, mois 13-15), avec le concours du/ de la post-doc nouvellement recruté.e. S'ensuivront : la rédaction et la soumission du protocole expérimental au comité d'éthique (④① : mois 13-15) ; la mise en œuvre de tests pilotes par les mêmes (④② : mois 16) ; la phase d'expérimentation – tests et notation – en milieu scolaire pour le groupe 1 e et en laboratoire pour les deux groupes d'étudiants, prise en charge par le/la post-doc (④③ : mois 17- 22) ; la phase d'analyse des résultats par tous les membres de l'équipe (④④ : mois 23-26) ; la phase de valorisation, sur EcoLiRed (voir *supra* ②⑥, mois 35-42), et dans le cadre du réseau CANOPÉ, *via* des conférences et publications, co-effectuées et co-rédigées avec les membres de CLESTHIA, de STIH et Séverine Casalis, avec Ludivine Javourey-Drevet, pour les publics de l'Éducation nationale (④⑤ : mois 31-42). Les autres livrables (④⑥ : mois 26-42), concernant tous les scientifiques de SCALab, sont détaillés ci-dessous.

### Livrables

- ◆ Participation à l'ouvrage de synthèse de l'objectif 3.
- ◆ Publications dans des revues internationales disciplinaires (**Open accès**, voies dorée – *PlosOne, Journal of Cognition*– ou verte, avec *post prints* déposés immédiatement sur HAL ou LilloA pour les licences CC-BY) et conférences disciplinaires internationales.
- ◆ Publications prévues pour le/la post-doc : 3 articles (l'un sur les apprentis lecteurs niveau CM2 ; le deuxième sur la première étude menée auprès des étudiants ; le troisième sur la deuxième étude menée auprès des étudiants en psychologie).
- ◆ Publication interdisciplinaire avec CLESTHIA – voir *supra*, Objectif 2, Livrables.

### Éthique et gestion des risques

**1. Éthique** – S'agissant de recueil de données comportementales chez l'enfant et l'adulte, le protocole sera soumis au comité éthique de Lille, et recevra un numéro de déclaration au registre des traitements de l'université (donné par le DPO). Dans la mesure où les données relevées ne sont pas identifiantes, une déclaration à la CNIL n'est pas prévue *a priori*. **2. Gestion des risques** – L'étude ne comporte pas de risque intrinsèque. La recherche de participants est parfois compliquée, mais l'équipe a tissé des réseaux auprès du

rectorat et de l'INSPE lui permettant de résoudre ces difficultés. Si les écoles devaient fermer à cause de la pandémie COVID, les expérimentations seraient décalées. Pour prévoir « large », les expérimentations en psycholinguistique démarreront dès le milieu de la 2<sup>e</sup> année du projet.

## II. Organisation et réalisation du projet

Le projet mobilisera deux équipes de **linguistique diachronique** – l'une, de **médiévistes / *Early Renaissance***, l'autre de **spécialistes du français préclassique et classique** – et une équipe de **psycholinguistes**.

### II.1 Coordinatrice scientifique

La coordination scientifique d'EcoLe sera assurée par Claire Badiou-Monferran (CBM), PR en sciences du langage à la Sorbonne Nouvelle, et rattachée à l'EA 7345 CLESTHIA, dont elle codirige un axe – l'axe « Linguistique de l'écrit ». Consacrés à la première modernité (16<sup>e</sup> -18<sup>e</sup> s.) mais aussi, plus largement, à l'histoire du changement linguistique pour toute la période du « français moderne » (16<sup>e</sup>-21<sup>e</sup> s.), les travaux de linguistique textuelle de CBM interrogent, dans une perspective diachronique, sur de gros corpus littéraires et non littéraires, l'interaction des codes oral et écrit. À ce titre, ils croisent les principaux observables du projet (phrase *vs* période ; ponctuation ; connecteurs). L'opération que CBM pilote actuellement sur la notion de « traces linguistiques et discursives » (<http://www.univ-paris3.fr/des-traces-linguistiques-et-discursives-en-sciences-du-langage-702353.kjsp>) devrait permettre de stabiliser le cadre théorique et méthodologique du projet.

Si CBM, avant EcoLe, n'avait encore jamais travaillé avec des psycholinguistes, et si cette dimension pluridisciplinaire est donc disruptive, au meilleur sens du terme, en revanche, dans son domaine de spécialité, elle a une bonne expérience des collaborations avec les linguistes contemporanéistes (parmi d'autres, Badiou-Monferran et Rossari 2016) et médiévistes (Badiou-Monferran et Marchello-Nizia 2020 ; Capin et Badiou-Monferran 2020). CBM consacrera 60% de son temps de recherche à EcoLe.

### II.2 Consortium

Le consortium réunit des personnes et des institutions aux compétences et à la complémentarité reconnues.

#### 2.1 Équipes de linguistique diachronique

##### 2.1.1 – Partenaire porteur : EA 4509 CLESTHIA, Sorbonne Nouvelle

Outre sa **compétence**, historique, en matière de **constitution de corpus** (corpus CFPP<sub>2000</sub>, FRACOV, Ecriscol), le laboratoire a développé, plus récemment, un programme de recherche en **diachronie longue** (voir parmi d'autres les opérations « Variation d'auteur et variation rédactionnelle : édition, génétique textuelle et e-philology » ; « Projet innovant *Lettres d'outre-mer, 1652-1760* »). Significativement, la réalisation du **CoDiF** (dir. Gabriella Parussa et Florence Lefeuve), corpus de textes dialogués dramatiques ou narratifs en vers et/ou en prose écrits entre le 13<sup>e</sup> et le 21<sup>e</sup> s., se situe à la croisée de ces deux orientations. Sans se superposer à EcoLe, le CoDiF sera en mesure de fournir au présent projet une partie de ses corpus.

Dans ce contexte, l'équipe réunie par Claire Badiou-Monferran autour du projet EcoLE est constituée :

- ◆ de **membres de CLESTHIA**, titulaires ou associé.e.s, toutes et tous spécialistes du français préclassique et classique mais ayant développé une expertise en amont (Olivier Halévy) ou en aval de la Première Modernité (Myriam Bergeron, Florence Lefeuve, Emily Lombardero, Ludovica Maggi, Anne-Régent Susini) ;
- ◆ de **membres du Réseau transnational RCFC** (Réseau corpus français préclassique et classique <https://wikis.fu-berlin.de/pages/viewpage.action?pageId=594411980>). Ces dernières, dont la spécialisation diachronique est large, et va également au-delà de la seule période (pré)classique, ont rejoint EcoLe en vue du partage des corpus et des outils des projets dont elles sont responsables :
  - projet BAILLY [correspondance politique, diplomatique, administrative et religieuse d'Albert Bailly. Lettres et gazettes, 1643-1688] porté par Antonella Amatuzzi (Turin), philologue, spécialiste de français préclassique et classique mais aussi de français médiéval ;
  - projet APWCF [lettres diplomatiques françaises échangées pendant le congrès préparatoire à la paix de Westphalie, 1643-1648] d'Annette Gerstenberg (Potsdam), sociolinguiste, spécialiste de la Première Modernité mais aussi de la période contemporaine. A. Gerstenberg est en effet à l'initiative du projet, purement synchronique, LangAge corpora, qui porte sur les enjeux linguistiques du vieillissement et qui inclut une réflexion sur les retombées de l'âge linguistique sur la lecture à voix haute (Gerstenberg et Kairet 2021) ;
  - projet SERMO [sermons protestants publiés à Genève, Paris, ou dans les pays du Refuge, entre 1552-1744, et transcrits diplomatiquement] par Carine Skupien Dekens (Neuchâtel), tout à la fois historienne de la langue et spécialiste de FLE.

**Ces profils diachroniques larges faciliteront les échanges avec les deux autres équipes** impliquées dans le projet : équipe STIH, réunissant des linguistes médiévistes et/ou spécialistes de l'*Early Renaissance* ; équipe SCALab, réunissant des contemporanéistes, psycholinguistes. Ils présentent en outre l'avantage de combiner **deux ordres de spécialité** :

[1] celle des **genres de discours** – discours religieux avec les sermons protestants (Skupien) *vs* catholiques (Régent) ; discours scientifique autour des encyclopédies et des dictionnaires (Amatuzzi, Badiou-Monferran) ; discours politique avec les correspondances diplomatiques (Amatuzzi ; Gerstenberg) ; discours littéraire (théâtre : Halévy, Maggi, Régent ; Fables et Contes : Badiou, Lombardero) ; discours familier (Lettres privées : Bergeron) ;

[2] celle des **observables** : ponctuation (Skupien, Badiou) ; marqueurs discursifs (Lefevre, Badiou) ; phrase *vs* période (les mêmes) ; vers *vs* prose (Halévy), dimension typo-iconographique de la textualité (Réach Ngô)... – voir ci-dessus I.b État de l'art.

Les séminaires, les JE, et ateliers transversaux (tâches ④ à ⑧) auront vocation à documenter les domaines non couverts (notamment, le domaine juridique, pourtant central avec le corpus de « cris publics ») *via* des échanges en interne avec les membres médiévistes de STIH (notamment, pour le même domaine : Hélène Biu) et/ou des échanges avec des expert.e.s extérieur.e.s au projet, spécialistes d'*Early Modern* (entre autres, Dinah Ribard, Anna Arzoumanov).

### 2.1.2 – STIH

Comme l'indique le nom développé « Sens, Texte, Informatique, Histoire », l'EA 4509 réunit des chercheurs spécialistes d'Histoire de langue et des chercheurs spécialistes de traitement informatique des langues (TAL).

L'équipe de **médiévistes** réuni.e.s autour de Gabriella Parussa, PR de linguistique et philologie à Sorbonne Université, dont les travaux de pragma-philologie sur les marqueurs discursifs et la ponctuation ont permis de documenter à nouveaux frais, pour la période médiévale, la question de l'inscription des marques d'oralité dans l'écrit, sera constituée :

- ◆ de collègues **membres de STIH** (Hélène Biu, Joëlle Ducos, Adeline Desbois-Ientile)
- ◆ de collègues **extérieur.e.s** (Andrea Valentini, CLESTHIA ; Elena Llamas Pombo, Salamanque, EMYRhd).

Toutes et tous sont spécialistes d'Ancien et Moyen Français et/ou de d'*Early Renaissance* (Ientile ; Parussa). Ils ont développé une expertise [1] dans divers domaines : juridique (Biu) ; scientifique (Ducos, Llamas Pombo) ; religieux (Parussa) ; littéraire (Parussa ; Desbois) et [2] sur des observables linguistiques différents (voir ci-dessus, I.b État de l'art). Ce chassé-croisé est la condition *sine qua non* d'un travail articulant les **dimensions évolutive** (paramètre temporel) et **situationnelle** (paramètre communicationnel) à une **dimension générique et textuelle** (genres de discours et types de textes).

### 2.1.3 – Encadrement transversal en humanités numériques et TAL

Par-delà le partenariat transcéculaire unissant CLESTHIA et STIH, la réalisation de la chaîne de traitement informatique et éditoriale visée par EcoLe doit pouvoir s'appuyer sur une double compétence : en Humanités numériques d'une part et en traitement automatique des langues de l'autre (TAL).

Pour ce qui concerne l'aspect **Humanités numériques** et **gestion de projets**, l'élaboration et la supervision de la chaîne de traitement des données seront confiées à Anne Réach-Ngô, dont les projets récents visent à articuler différentes phases de constitution et d'exploration de corpus – séminaire ITEM-TransLitterae « Corpus numériques et scénarios de recherche » avec Richard Walter (2019) ; Journées EVEille 2021 « Constitution des données de la recherche » et EVEille 2022 « Écosophie du numérique. Mettre en œuvre les principes FAIR(e) » (UHA) ; projet de constitution d'une chaîne éditoriale en partenariat avec la MISHA en 2022-2023.

Les travaux informatiques sur la plateforme EMAN et l'adaptation des outils conçus pour le projet seront confiés à Richard Walter, responsable de la plateforme.

La **composante TAL** (transcription automatique HTR et OCR, annotation automatique de corpus) sera pour sa part supervisée par Gaël Lejeune (STIH), directeur adjoint du CERES (le Centre d'expérimentation en méthodes numériques pour les recherches en sciences humaines et sociales). Les travaux de G. Lejeune (GL) documentent l'impact de la variation des données (principalement mais sans exclusive : multilinguisme, hétérogénéité, bruitage) sur l'applicabilité des méthodes de traitement automatique des langues. Du fait de son implication dans le projet Antonomaz (ANalyse auTOMatique et NumérisatiON des MAZarinades, dir. Karine Abiven), GL a développé une bonne expertise de la diachronie, tout particulièrement des corpus d'ancien régime.

Pour l'utilisation des serveurs informatiques, la finalisation de « package » logiciels, ou encore l'assistance technique, EcoLe bénéficiera de **l'appui de l'unité de service CERES**.

Dans le cadre du présent projet, Gaël Lejeune (déjà membre de STIH) et son équipe (principalement, un.e Post-doc de 24 mois financé.e par l'ANR) seront rattachés à STIH. Anne Réach-Ngô (ITEM), Richard Walter (THALIM, Sorbonne Nouvelle) et leur équipe (principalement, 2 ingénieur.e.s d'études recruté.e.s pour 12 mois chacun, sur financement de l'ANR) seront rattachés à CLESTHIA. Les deux équipes feront dialoguer gestion de projet éditoriale et numérique avec les besoins en traitements linguistiques du projet.

## 2.2 L'équipe de psycholinguistes (SCALab, Lille)

L'**UMR 9193 SCALAB** de l'Université de Lille est un laboratoire situé à la frontière des SHS et des neurosciences. En son sein, Séverine Casalis – coresponsable de l'équipe « Langage », chevalier dans l'ordre des palmes académiques depuis novembre 2021 – a développé un ensemble de travaux sur la lecture et sur la dyslexie (<https://scholar.google.fr/citations?hl=fr&user=kOGLEUQAAAAJ>). Pour EcoLe, elle collaborera avec Ludvine Javourey-Drevet (SCALab), spécialiste des questions de fluidité et de compréhension en lecture ; Gwendoline Mahé (SCALab), dont la dyslexie constitue le domaine d'expertise ; Laurent Sparrow (SCALab), spécialiste des phénomènes attentionnels impliqués en lecture, et ayant développé une compétence particulière en pupillométrie (*Eye Movements*). L'équipe prévoit d'être épaulée par un.e Post-doc de 18 mois (sur un financement ANR), et par des vacances (voir ci-dessous).

### II.3 Moyens mis en œuvre et demandés pour atteindre les objectifs

Projet ANR EcoLe	CLESTHIA	STIH	SCALAB
Frais de personnel	194 600 1 contrat doctoral 2 IE Vacations	107 200 1 post-doc (24 mois)  Vacations	77 400 1 post-doc (18 mois)  Vacations
Instruments et matériel	3000	1000	1000
Prestations de service	18 000	22 000	3000
Frais généraux	15 500	10 500	7500
TOTAL DÉPENSES	233 100	140 700	93 700
Frais de gestion : + 13%	30 303	18 291	12 181
SOUS-TOTAL	<b>263 403</b>	<b>156 731</b>	<b>101 340, 66</b>
Aide allouée	<b>521 474,66 euros</b>		

**CLESTHIA – Frais de personnel** – EcoLe est un projet exigeant en personnel, en raison de la chaîne de traitement informatique et éditoriale qu'il engage. CLESTHIA prendra en charge la composante « Humanités numériques », soit : [1] **2 CDD d'ingénieurs d'études de 12 mois chacun** [mois 2-13 ; 24- 35] venant servir de renfort à R. Walter, l'IR responsable de la plateforme EMAN, en début et en fin de projet, pour l'ensemble des tâches détaillées sous I.c (2 x 40000 euros) ; [2] **4 mois de vacances** (9600 euros), pour la révision semi-diplomatique de tout ou partie des corpus des sites partenaires (Bailly et APWCF), ayant vocation à intégrer EcoLiReD. À quoi s'ajoute [3] la demande d'un **contrat doctoral** financé pour trois ans (105 000 euros). La tâche de ce dernier consistera à redoubler la documentation existante sur les pratiques de lecture en partant, non pas de ces dernières mais des « imaginaires, représentations et des pratiques du texte, du 13<sup>e</sup> au 18<sup>e</sup> siècle », et en privilégiant des sources peu exploitées (traités de grammaire, ouvrages de rhétorique, manuels d'éloquence, remarques, dictionnaires, etc.). La/le doctorant.e contribuera, par cette moisson de témoignages externes inédits, à la stabilisation des différents corpus du projet. Il/elle sera associé.e aux travaux d'annotation de ces derniers. Enfin, il/elle sera partie prenante de l'organisation des JE, ateliers, séances de séminaire, à la préparation/organisation desquels il/ elle aidera. Il/elle sera associé.e à la valorisation du site EcoLiReD et à l'écriture de certaines publications. **Coût en matériel** – un ordinateur pour chacun.e des personnels recrutés (les 2 IE et le/la doctorant.e : 3000 euros). **Coût des prestations de service** – L'achèvement de la numérisation du corpus de français préclassique et classique est évalué à 4000 euros (0,5 euro la vue). Sa transcription automatique HTR et OCR à 6000 euros. CLESTHIA contribuera par ailleurs pour moitié (5000 euros) au développement du site EcoLiReD (nécessitant l'installation de nouvelles fonctionnalités sur EMAN – intégration des outils textométriques) et pour un tiers (3000 euros) à sa valorisation (prise en charge notamment de l'élaboration de livrets pédagogiques à destination des enseignants). **Frais généraux additionnels** – CLESTHIA prendra en charge l'organisation du séminaire annuel, de la réunion de coup d'envoi (kick off), et d'un atelier de travail intermédiaire (le tout pour la somme de 9000 euros), ainsi que du colloque de fin de projet (3 jours : 5000 euros). Elle réservera la somme de 2700 euros aux missions de valorisation des résultats (conférences, colloques) et celle de 800 euros à une aide à la publication en *gold open access*.

**STIH – Frais de personnel** – STIH prendra en charge la composante TAL du projet EcoLe. L'équipe recrutera un.e post-doc de 24 mois (mois 18-41), en charge de l'annotation des corpus (100 000 euros). Elle veillera à intégrer ce.tte dernier.e, confronté.e à de nouvelles questions liées à la complexité de son objet (annotation d'un texte dans sa globalité, dans une perspective de linguistique textuelle), aux travaux de valorisation (colloques, conférences, publications). Des vacances (3 mois : 7200 euros) viendront servir de renfort à l'annotation des corpus d'une part, d'autre part, au travail d'intégration des corpus sur le site, et enfin, en fin de projet, aux ajustements de son interface après les premiers retours d'utilisateurs. **Coût en matériel** – Achat d'un ordinateur pour la/le post-doc. **Coût des prestations de service** – L'achèvement de la numérisation du corpus de français médiéval et *Early Renaissance* est évalué à 5000 euros (0,5 euro la vue). Sa transcription automatique HTR et OCR à 9000 euros. STIH contribuera par ailleurs pour moitié (5000 euros) au développement du site EcoLiReD (nécessitant l'installation de nouvelles fonctionnalités sur EMAN – intégration des outils textométriques) et pour un tiers (3000 euros) à la valorisation de ce dernier (via la mise en place d'une exposition virtuelle de manuscrits, incunables et imprimés présentant des caractéristiques de mise en page spécifiques ; et via des enregistrements de lecture à voix haute). **Frais généraux additionnels** – STIH participera à l'organisation d'un atelier de travail et d'une JE intermédiaires (5000 euros). L'équipe réservera 2700 euros à ses missions de valorisation (conférences, colloques) et 800 euros à une aide à la publication en *gold open access*.

**SCALab – Frais de personnel** – Pour les travaux d'administration des tests (en milieu scolaire pour les CM2 et en laboratoire pour les lecteurs experts) et des travaux de cotation, SCALab recrutera un.e post-doc sur 18 mois (Mois 13-28 : 75 000 euros). L'équipe associera ce.tte dernier.e à l'élaboration des tests de compréhension, à l'analyse des résultats et à la rédaction des articles scientifiques. **Temps de travail pour le recueil et la cotation des données** – Le nombre minimal de participants sera fixé à 50 pour les groupes



1 et 2 (CM2 ; étudiants). Partant, le temps total de passation s'élèvera à 475 heures (CM2 : 15 minutes par texte soit 15 x 12 textes pour la lecture + 30 minutes de mesures générales – lecture, vocabulaire, niveau cognitif non verbal – x 50 participants = 175 heures de passation ; Adultes : 3 heures x 50 x 2 = 300 heures de passation). À cela viendront s'ajouter 200 heures de cotation : soit, un total de 675 heures de travail expérimental. 1 vacation (1 mois : 2400 euros) suppléer le/la post-doc dans son travail de collecte des données, très long. **Coût des prestations de service** – SCALab rendra disponible sur EcoLired des jeux interactifs en lien avec la lecture à haute voix. Jeux dont l'élaboration nécessitera quelques prestations de service (coût : 3000 euros). **Frais généraux additionnels** – L'équipe organisera la JE intermédiaire de méthodologie sur l'interdisciplinarité (JE de 2 jours : 4000 euros). Elle réservera 2700 euros aux missions de diffusion/valorisation (colloques, conférence) et 800 euros à une aide à la publication en *gold open access*.

### III – Impact et retombées du projet

L'impact d'EcoLe est tout à fois scientifique et sociétal.

#### A/ Impact scientifique

**1. Production d'un nouveau savoir** – En soi, l'idée d'une **grammaire des pratiques de lecture**, conditionnant le succès *vs* insuccès de chacune d'entre elles, est proprement disruptive. Le présent projet, qui se propose de l'étayer en faisant retour sur une période phare, conduisant du 13<sup>e</sup> au 18<sup>e</sup> siècle, aboutira à la **production d'un nouveau savoir** dans chacun des deux domaines impliqués par le projet : les sciences de la culture d'une part, la psycholinguistique de l'autre. **1.1 Impact au sein des sciences de la culture** – En invitant dans le débat, de manière frontale, la question des *genres de discours*, et en faisant l'hypothèse que suivant les genres discursifs, les pratiques de lecture ont évolué à des *vitesse*s et à des *moments différents*, EcoLe vise à une **description intégrée de l'histoire de la lecture**, combinant tout à la fois les acquis de l'approche « évolutive » et ceux de l'approche « variationniste ». Sa modélisation, dynamique et non linéaire, documentant des phénomènes d'émergence précoce et/ou de rémanence (au sens que les participants ont donné à ces termes dans des travaux antérieurs : voir Badiou-Monferran et Ducos, dir., 2012 ; ou encore Badiou-Monferran, Parussa et Llamas Pombo dans Badiou-Monferran, dir ; 2020) se propose de renouveler l'histoire de la lecture, en en instruisant une représentation à géométrie variable, de type modulaire. **1.2. Impact en psycholinguistique** – Les modèles d'apprentissage de la lecture n'intègrent ni la question de la modalité de lecture, ni celle de l'adéquation entre architecture textuelle et modalité de lecture. Les travaux menés permettront d'ouvrir une question nouvelle, très pertinente pour la lecture (et la question de la spécificité des codes phonologiques activés silencieusement *vs* à voix haute). En outre, les travaux permettront de relier deux champs de recherche, celui de la compréhension en fonction de la modalité et celui, plus développé (Rouet, 2006) du rôle de l'architecture textuelle sur la compréhension et la recherche d'information. Ces deux questions, posées de façon totalement indépendante, se trouveront interconnectées dans une perspective totalement innovante de l'étude cognitive de la lecture.

**2. Production d'une méthode non conventionnelle** – Partant du constat que très nombreux sont encore aujourd'hui les genres de discours et les types de textes dont on ne sait absolument pas comment ils pouvaient être lus (voir *supra* les collections – dite « flottantes » – de textes non assignés), le présent projet propose de compléter la documentation des historiens de la lecture, qui repose sur la collecte de témoignages – prescriptifs ou descriptifs –, par une **documentation probabiliste**, mobilisant un **point de vue interne**. Supposant qu'indépendamment de toute instruction de lecture explicite, les textes ont vocation à encoder, via une série de choix linguistiques et typo-iconographiques, le mode d'appropriation (par l'œil ou par l'oreille) auquel ils sont préférentiellement destinés, EcoLe entend reconstruire, à partir d'**indices textuels**, pour les écrits du passé, des **pratiques de lecture probables**. La méthodologie requise en appelle donc au concept de « **trace** ». Soit : à un **mode de connaissance indirect**, inférant à partir des effets. Si cette démarche est devenue centrale en SHS, après les travaux de Derrida (1972), Ricoeur (2000) ou encore Ginzburg (1989, « Traces, racines d'un paradigme indiciaire »), elle demeure peu exploitée en sciences du langage (voir ci-dessus, II.1, section « Coordinatrice »). EcoLe entend démontrer (tout du moins, éprouver) son opérativité pour la linguistique diachronique, en la mettant à l'épreuve de la linguistique de corpus, et en la confrontant aux acquis de la linguistique textuelle.

**3. Extension du champ des collaborations interdisciplinaires** – La collaboration de linguistes diachroniciens et de psycholinguistes est très largement inédite. Dans le cadre du présent projet, elle était rendue doublement nécessaire (voir *supra*). On peut préjuger deux niveaux de rendement. **3.1 Rendement faible** – Les séminaires et JE auront permis de construire une culture commune, un langage commun, une méthodologie commune : autrement dit, d'importer dans chaque champ disciplinaire impliqué des concepts, des outils et des questionnements exogènes. **3.2 Rendement fort** – L'hypothèse collaborative qu'EcoLe entend tester est que (i) les expérimentations psycholinguistes doivent pouvoir constituer une instance de validation empirique pour la description et l'interprétation des phénomènes linguistiques du passé (ii) que, réciproquement, les états de langue passés doivent pouvoir fournir aux psycholinguistes des observatoires inédits. Le cas échéant, le présent projet ouvrira la voie d'une collaboration interdisciplinaire encore jamais tentée, contribuant à l'émergence d'une discipline nouvelle : la **psycholinguistique diachronique**.

#### B/ Impact sociétal

À l'heure où l'époque contemporaine, héritière d'une longue tradition tabulaire et visuelle, redécouvre le bénéfice de la lecture « à voix haute » dans différents domaines :

- celui du **milieu scolaire**, dont les instructions ministérielles recommandent, depuis 1992, d'« augmenter le temps dévolu à la lecture à voix haute, en vue d'améliorer la fluence et la compréhension en lecture » (niveau CE1-CM2) ; ou encore, d'« aborder les textes à partir de lectures à haute voix [...] pour travailler l'attention et la concentration et développer les représentations mentales » (niveau 6<sup>e</sup>-4<sup>e</sup>) (MEN 2020) ;

- celui du **milieu paramédical (orthophonie)**, où différents travaux ont montré les « effets positifs » de la lecture à voix haute sur diverses pathologies, dont la « dyslexie » (Regnier 2010) ;
- celui du **milieu culturel**, où se multiplient les lectures en « tiers-lieux » – sur la scène, les places publiques, dans les bibliothèques, les résidences –, visant l’élaboration de nouvelles représentations collectives, et constituant un mode de socialisation « quasi-religieux », (Dheur 2017 ; ou encore Petit 2014) ;
- celui, **privé**, de la **sphère familiale**, où la lecture à haute voix engage un « mode de partage transitionnel » entre l’enfant et l’adulte (Merlin-Kajman 2016),

la question du rôle que peuvent jouer – ou non – les différents formats architecturaux dans le succès *vs* insuccès du passage d’une pratique de lecture à une autre est devenue un **enjeu sociétal**. Si le degré de compréhension (pour le lecteur) des écrits lus à haute voix est conditionné par la nature de la « mise en page » et de la « mise en texte », autrement dit, s’il existe des architectures textuelles plus adaptées que d’autres à telle ou telle pratique de lecture, alors les corpus institués (canon scolaire, médical, culturel, de littérature jeunesse) seraient sans doute à repenser. C’est tout le propos de Baetens (2016) qui, dans un ouvrage pionnier, retrace l’histoire des mauvaises pratiques de lecture de la poésie moderne et contemporaine : « Dans bien des cas », explique-t-il, « un poème est diminué par le saut de la page à la scène », notamment s’il relève d’un régime « grammatextuel » ou « bruitiste », (Baetens 2016 : 9 et suiv.). Par-delà les « ratés » culturels, qu’en est-il, plus largement, des instructions de lecture en régime scolaire ? Le format du texte suivant, emprunté à un « Livret d’entraînement à la lecture fluide » :

Je lis à haute voix en respectant la ponctuation :

“ Ce matin, un employé municipal s’est garé devant mon immeuble. Il a pris une pelle et a fait un grand trou, là, juste devant chez moi.

C’est sûr, il va revenir pour planter un arbre.

C’est une très bonne nouvelle.” (*Livret[s] d’entraînement à la lecture fluide*, niveau CE1, Hatier, 2019, p. 3)

est-il adapté à la consigne de lecture “ à haute voix ” ? Et que penser des formats de la collection « Dyscool » (Nathan), collection dite de « lecture facilitée », à destination des « enfants dyslexiques et [de] tous ceux qui ont du mal à entrer dans la lecture » (Dyscool, p. 2)? EcoLe se propose de sensibiliser l’Éducation nationale et le milieu enseignant à cette question :

- ◆ en construisant une offre de formation sur le réseau CANOPÉ [cours en ligne ; workshop]
- ◆ en donnant des conférences pédagogiques dans les INSPE
- ◆ en soumettant des articles de diffusion dans des revues pour professionnels
- ◆ en ouvrant le site EcoLiReD au public enseignant et aux scolaires dès la troisième année (tests de l’interface ; de la lisibilité des livrets pédagogiques déposés ; de la fonctionnalité des jeux interactifs)
- ◆ en donnant des entretiens dans les Médias assurant la promotion des lectures à haute voix (plateforme éducative numérique de l’audiovisuel public français LuMi. <https://www.lumni.fr/> ; France Télévisions, *La Grande Librairie*, concours « Si on lisait à voix haute », depuis 2020)
- ◆ en développant une offre de formation en amont, dans le cadre des Masters MEEF.

La plupart des chercheurs impliqués dans EcoLe sont également des enseignants. Une grande partie participe à la formation initiale et/ou continue des enseignants dans diverses académies. Les relations entretenues depuis de nombreuses années avec les corps d’inspection et les formateurs de terrain, les réseaux tissés par l’équipe de psycholinguistes auprès des Rectorat et INSPE, faciliteront la diffusion des résultats à tous les niveaux décisionnels.

Ainsi, un enjeu majeur est d’éclairer, à l’aide d’une vision pluridisciplinaire, les décideurs en matière d’enseignement scolaire et de rééducation. Par ailleurs, le projet permettra également l’accès au grand public à des textes anciens, parfois adaptés à l’esprit contemporain. Cette connaissance du passé a aussi sans conteste un impact sociétal fort, permettant à différents acteurs de s’emparer d’éléments de la culture parfois difficiles d’accès. Le grand public sera également sensibilisé à l’intérêt de la lecture à voix haute via le site et des actions de médiation scientifiques (jeux, conférences grand public, médias, etc).

## IV. Bibliographie

[Noms des participants **en gras**]

ABIVEN K. et **LEJEUNE G.**, 2019. Analyse automatique de documents anciens : tirer parti d’un corpus incomplet, hétérogène et bruité. London: ISTE,

[https://www.openscience.fr/IMG/pdf/iste\\_ridows18v2n1\\_3.pdf](https://www.openscience.fr/IMG/pdf/iste_ridows18v2n1_3.pdf)

ADAM J.-M., 2018. *Le Paragraphe : entre phrases et texte*. Paris, A. Colin.

**AMATUZZI A.**, **AYRES-BENNETT W.**,

**GERSTENBERG A.**, **SCHÖSLER L.** et **SKUPIEN DEKENS C.**, 2019.

Améliorer et appliquer les outils numériques. Ressources et approches pour l’étude du changement linguistique en français préclassique et classique. *Le français en diachronie*. Strasbourg, ELiPhi, 337-364.

ARABYAN M., 1994. *Le paragraphe narratif*. Paris, L’Harmattan.

— 2012. *Des lettres de l’alphabet à l’image du texte. Recherches sur l’énonciation écrite*. Limoges, Lambert-Lucas.

**BADIOU-MONFERRAN C.**, 2020. Sémantique des coordonnants ; Avec Ch. Marchello-Nizia, Sémantique des connecteurs. *Grande Grammaire Historique du français*. Bruxelles, De Gruyter Mouton, 1633-1674.

— éd., 2020. *La rémanence : un concept opératoire pour la linguistique diachronique ? Le cas du français*, in *Le français moderne*, 2020/2.

— 2019. Segmentation des discours (pré-)classiques et modèle périodique fribourgeois. *Phrase et période entre les XV<sup>e</sup> et XVIII<sup>e</sup> siècles*, *Verbum* XLI, 273-301.

— 2018. De l’écrit oralisé à l’oral scripturalisé. L’évolution des

- emplois de *et*, jonctif de phrases et de propositions, dans les fictions narratives en prose des 16<sup>e</sup> et 17<sup>e</sup> siècles. *L'oralité dans le roman (XVI<sup>e</sup> et XVII<sup>e</sup> siècles)*, *Elseneur* 32, 53-74.
- 2017. Enregistrement(s) de la parole et lecture(s) de l'écrit. La ponctuation de *La Barbe bleue* dans les éditions lettrées et peu lettrées des XVII<sup>e</sup> XVIII<sup>e</sup> et XIX<sup>e</sup> siècles. *La mise à l'écrit et ses conséquences*. Tübingen, A. Francke Verlag Tübingen, 189-210.
- 2016. Avec C. Rossari. « *Donc* et la complétion du discours », *Le français moderne*, 2016/2, 148-163.
- 2012. Avec J. DUCOS, éd. *L'émergence : un concept opératoire pour les sciences du langage ?* in *L'Information grammaticale*, n° 134.
- BAETENS, J., 2016. *À voix haute. Poésie et lecture publique*. Bruxelles, Les Impressions Nouvelles.
- BEAUME E., 1987. La lecture à haute voix. *Les actes de lecture* 18. [https://www.lecture.org/revues\\_livres/actes\\_lectures/AL/AL18/AL18P14.pdf](https://www.lecture.org/revues_livres/actes_lectures/AL/AL18/AL18P14.pdf)
- BERGERON-MAGUIRE M.** et GREUB Y., 2020. Qui n'entend qu'une cloche n'entend qu'un son. Le français classique des lettres interceptées durant les guerres de course atlantiques. En ligne, *CMLF* 2020.
- BERNINGER V. W. & ABBOTT R. D., 2010. Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in grades 1, 3, 5, and 7. *Journal of educational psychology*, 102(3), 635.
- BIBER D., 2015. Corpus-Based and Corpus-driven Analyses of Language Variation and Use. *The Oxford Handbook of Linguistic Analysis* (2 ed.), 1-32.
- BIET Ch., 2006. L'oral et l'écrit. *Histoire de la France littéraire*, t. 2, Paris, PUF, 409-434.
- BOUCHET F., 2008. *Le discours sur la lecture en France aux XIV<sup>e</sup> et XV<sup>e</sup> siècles : pratiques, poétiques, imaginaires*. Paris, Champion.
- CAPIN D. et **BADIOU-MONFERRAN C.**, 2020. À l'aune de la transcatégorisation : modélisation des emplois de *ET* dans la langue d'hier et d'aujourd'hui. *Bulletin de la Société de Linguistique de Paris*, p. 134-193.
- CASALIS S.**, 2019. Avec R. KHELIFI et **L. SPARROW**. Is a Frequency Effect Observed in Eye Movements During Text Reading? A Comparison Between Developing and Expert Readers. *Scientific Studies of Reading*, Taylor & Francis (Routledge), 334-347.
- 2013. Avec C. LEUWERS et H. HILTON. Syntactic comprehension in reading and listening: a study in French dyslexic children. *Journal of Learning Disabilities* 46, 201-219.
- 1996. Avec P. LECOCQ, C. LEUWERS et N. WATTEAU. *Apprentissage de la lecture et compréhension d'énoncés*. Lille, Presses Univ. du Septentrion.
- CAVALLO G. et CHARTIER R., [1997] 2000. *Histoire de la lecture dans le monde occidental*, Paris, Seuil.
- CHAYTOR H.J., 1945. *From Script to Print: An Introduction to Medieval Vernacular Literature*. Cambridge, Heffer and sons.
- COLEMAN J., 1996. *Public Reading and the Reading Public in Late Medieval England and France*. Cambridge, CUP.
- CONQUET A. et RICHAUDEAU, F., 1973. Cinq méthodes de mesure de la lisibilité, *Communication et langages* 17, 5-16.
- DARNTON R., 1971. Reading, Writing and Publishing in 18<sup>th</sup>-Century France: A Case Study in the Sociology of Literature. *Daedalus*, 214-256.
- DEMONET M.-L., 2016. Rhétorique de l'imprimé à la Renaissance. *HEL* 9, 146-161.
- DESBOIS-IENTILE A.**, 2019. Ponctuer la prose à la Renaissance. L'exemple de Lemaire de Belges (1511-1549). *Verbum* XLI, 159-183.
- DHEUR S., 2017. « La lecture à voix haute. Entre écriture et oralité, une autorité en jeu », *BBF*, 168-185.
- DOUCET A. *et al.*, 2013. Overview of the ICDAR 2013. Competition on Book Structure Extraction. *12<sup>th</sup> International Conference on Document Analysis and Recognition (ICDAR<sub>2013</sub>)*. Washington, 1438-1443.
- DUCOS J.** 2019. De la glose à la parenthèse. Traduire Végèce en français au XVI<sup>e</sup> siècle. *Transmission of Knowledge in the Late Middle Ages and the Renaissance*. Turnhout, Brepols, 169-183.
- DUMONCEAUX P., 1990. La lecture à haute voix des œuvres littéraires au XVII<sup>e</sup> siècle, *Littératures classiques* 12, 117-125.
- FAUGERE M., à paraître. *Cette lecture qu'on ne peut connaître. Représentations de la lecture et fables du lecteur au XVII<sup>e</sup> siècle*. Doctorat de la Sorbonne Nouvelle, dir. H. Merlin-Kajman, soutenu en 2020.
- FUMAROLI M., et GOEME Ch., 1990. La parole vive au XVII<sup>e</sup> siècle, *Littératures classiques* 12, 7-11.
- GARCIA-RODICIO *et al.*, 2018. A comparison of reading aloud, silent reading and follower reading. Which is best for comprehension?/Una comparación de lectura en voz alta, lectura en silencio y lectura de seguimiento. ¿Cuál es mejor para la comprensión?. *Infancia y Aprendizaje*, 41(1), 138-164.
- GERSTENBERG A.** et KAIRET J., 2021. Prosodic Features and Situational Settings: Doing Reading Aloud in a French Writing Class. *Studia romanica et linguistica*, 141-156.
- GIGUET E. et **LEJEUNE G.**, 2019. Daniel@FinTOC-2019 Shared Task: TOC extraction and Title Detection, *Proceedings of the Second Financial Narrative Processing Workshop (FNP<sub>2019</sub>)*.
- GILMONT J.-F., 2003. *Le livre et ses secrets*. Genève, Droz.
- GOLDMAN, S. R., & RAKESTRAW J. A. Jr, 2000. Structural aspects of constructing meaning from text. *Handbook of reading research*, vol. 3, 11-335.
- HALE A. D. *et al.* 2011. An investigation of silent versus aloud reading comprehension of elementary students using maze assessment procedures. *Psychology in the Schools*, 48(1), 4-13.
- HALE A. D *et al.*, 2007. Comparing comprehension following silent and aloud reading across elementary and secondary students: Implication for curriculum-based measurement. *The Behavior Analyst Today*, 8(1), 9.
- HALEVY H.**, 2011. De la ponctuation manuscrite à la ponctuation imprimée : l'exemple de *l'Electra* de Lazare de Baïf. *La Ponctuation à la Renaissance*. Paris, Classiques Garnier, 71-92.
- HAUG H., 2014. Relecture critique de l'histoire de la lecture. Régularités discursives chez les historiens modernes. *Le Moyen Âge* 120, 123-133.
- HOOVER W. A. & GOUGH P. B., 1990. The simple view of reading. *Reading and writing*, 2(2), 127-160.
- JUGE R., BENTABET I. et FERRADANS S., 2019. The fintoc-2019 shared task: Financial document structure extraction. *Proceedings of the Second Financial Narrative Processing Workshop (FNP<sub>2019</sub>)*.
- KRIEBER M. *et al.* 2017. Eye movements during silent and oral reading in a regular orthography: Basic characteristics and correlations with childhood cognitive abilities and adolescent reading skills. *PLoS one*, 12(2), e0170986.
- KRIPPENDORFF K. *et al.*, 2016. On the reliability of unitizing textual continua: Further developments. *Quality & Quantity* 50.6, 2347-2364.
- LECOCQ, P. (1996). *L'É. co. s. se une épreuve de compréhension syntactico-sémantique (manuel et épreuve)*, 2 vol. Lille, Presses Univ. Septentrion.
- LEFAVRAIS P., [1967] 2005. *L'Alouette-R*. Centre de psychologie appliquée.
- LEFEUVRE F.**, 2014. *Étude grammaticale du français classique dans les textes*. Paris, Presses de la Sorbonne Nouvelle.
- LLAMAS POMBO E.**, 2020. La ponctuation. *GGHF*. Bruxelles, De Gruyter Mouton, 592-614.
- LOMBARDEO É.**, à paraître sur le site *Mouvement-Transitions*. Le féminin dans les *Contes de fées* d'Aulnoy : perspectives diachroniques sur la langue inclusive.
- 2018. La segmentation des romans et des nouvelles au XVII<sup>e</sup> siècle : théories et pratiques de la fiction. *CMLF<sub>2018</sub>*. <https://hal.archives-ouvertes.fr/hal-02456245/document>

- 2017. Formes et enjeux de l'exophore mémorielle dans les *Contes* de La Fontaine. *Seuils du nom propre*. Limoges, Lambert Lucas, 35-56.
- MAGGI L.**, 2019. *Herméneutique, oralité, temporalité. L'écriture traductive théâtrale de l'interprétation des classiques à la mise en voix*. Phèdre et Dom Juan traduits pour la scène italienne contemporaine. Doctorat soutenu en 2019 à la Sorbonne Nouvelle, dir. C. Laplace. <https://tel.archives-ouvertes.fr/tel-02165937>
- MANESSE D. et SIOUFFI G., eds, 2019. *Le féminin et le masculin dans la langue. L'écriture inclusive en questions*. Paris, ESF Sciences humaines.
- MARTIN H. J., éd. 2000. *La Naissance du livre moderne (XIV<sup>e</sup>-XVII<sup>e</sup> siècles). Mise en page, et mise en texte du livre français*. Paris, éd. du Cercle de la Librairie.
- MARTIN H. J. et CHARTIER R., 1982-1984. *Histoire de l'édition française*, t. 1 et 2, Paris, Promodis.
- MASSONNIE J., BIANCO M., LIMA L. & BRESSOUX P., 2019. Longitudinal predictors of reading comprehension in French at first grade: Unpacking the oral comprehension component of the simple view. *Learning and Instruction*, 60, 166-179.
- MAUREL F. et al., 2003. Oralisation des structures visuelles : de la lexico-syntaxe à la prosodie. *Interfaces prosodiques* 2003, 136-142.
- MESCHONNIC H., 1998. Tout ce qu'on ne sait pas qu'on entend. *À haute voix. Diction et prononciation aux XVI<sup>e</sup> et XVII<sup>e</sup> s.* Paris, Klincksieck, 223-235.
- MERLIN-KAJMAN H., 2016. *Lire dans la gueule du loup. Essai sur une zone à défendre, la littérature*. Paris, Gallimard, coll. « NRF ».
- MOIRAND S., 2021. L'écriture inclusive entre langue, discours et société. <https://www.youtube.com/watch?v=9ujVRx8lYyM>
- NOILLE-CLAUZADE Ch., 2004. Le sillon et la liaison : rémanence d'un modèle scriptural en rhétorique. *Littératures classiques* 50, 101-118.
- O'BRIEN B. A., WALLOT S., HAUSSMANN A., & KLOOS H., 2014. Using complexity metrics to assess silent reading fluency: a cross-sectional study comparing oral and silent reading. *Scientific Studies of Reading*, 18(4), 235-254.
- ONG, W. J., [1982] trad. fr. 2014. *Oralité et écriture*. Paris, Les Belles Lettres.
- OUELLETTE G. & BEERS A., 2010. A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and writing*, 23(2), 189-208.
- PARUSSA G.**, éd., 2020. Partie 4 de la GGHF. Codes de l'écrit : graphies et ponctuation ». GGHF. Bruxelles, De Guyter Mouton, 491-614.
- 2020. Avec **F. LEFEUVRE**, eds. *L'oral représenté en diachronie et en synchronie : une voix d'accès à l'oral spontané ? Langages* 217.
- 2018. La représentation de l'oral à l'écrit et la diachronie du français : un nouveau projet de recherche. *Nouvelles voies d'accès au changement linguistique*. Paris, Classiques Garnier, 181-199.
- 2017. Avec M. COLOMBO TIMELLI et **E. LLAMAS POMBO**, eds. *Enregistrer la parole et écrire la langue dans la diachronie du français*. Tübingen, Narr, col. ScriptOralia 143.
- PASCUAL E., 1996. Integrating Text Formatting and Text Generation. *Trends in Natural Language Generation: an Artificial Intelligence Perspective*. Springer Verlag, 205-221.
- PETTIT M., 2014. *Lire le monde : expérience de la transmission culturelle aujourd'hui*, Paris, Belin.
- PLED B., ROUDY P. et HAMEAU C., 1997. *Lire à haute voix au cycle 3*. Paris, Nathan pédagogie.
- POTOCKI A. et BILLOTTET É. 2020. *Numérique et apprentissages scolaires. Incidences du numérique sur l'apprentissage du lire, dire, écrire*. CNAM/CRESCO. <https://hal-cnam.archives-ouvertes.fr/hal-03234543>
- RABATEL A. et ROSIER L., eds, 2019. *Les défis de l'écriture inclusive. In Le discours et la langue. Revue de linguistique française et d'analyse du discours*, Tome 11.1.
- RAVEN J. C. & COURT JH, 1938. *Raven's progressive matrices*. Los Angeles, CA: Western Psychological Services.
- REACH-NGO A.** et al (eds), 2011. *Le discours du livre. Mise en scène du texte et fabrique de l'œuvre sous l'Ancien Régime*. Paris, Classiques Garnier.
- 2018. Avec **R. WALTER**. *Les Trésors de la Renaissance. La production d'une bibliothèque numérique sous Omeka envisagée comme outil de recherche. Séminaire "Génétique éditoriale de la première modernité"*, ITEM ([halshs-02883092](https://halshs-02883092))
- REGENT-SUSINI A.**, 2017. De l'oral à l'écrit : les Sermons du P. Coton mis en forme de méditations, ou de l'usage du paragraphe dans le livre de dévotion. *Papers on French Seventeenth Century Literature*, 321-336.
- REGNIER F., 2017. *Apports de la méthode « Voix haute » chez l'enfant dyslexique*. <https://dumas.ccsd.cnrs.fr/dumas-01523611>
- ROS-DUPONT M., éd., 1999. *La lecture à haute voix : du CP au CM2*. Paris, Bordas.
- ROUET J.-F., 2006. *The skills of document use: From text comprehension to Web-based learning*. London, Routledge.
- ROY M., GAONACH D. et FAYOL M., 2002. L'effet de la ponctuation et des connecteurs sur le traitement des phrases en lecture. *Le Langage de l'homme* 37, 139-161.
- SAENGER P., 1997. *Space Between Words: The Origins of Silent Reading*. Stanford, SUP
- SAUVAL K., PERRE, L. & CASALIS S., 2017. Automatic activation of phonological code during visual word recognition in children: a masked priming study in grades 3 and 5. *Reading and Writing*, 30(1), 51-67.
- SHARE D. L., 1995. Phonological recoding and self-teaching: *Sine qua non* of reading acquisition. *Cognition*, 55(2), 151-218.
- SIOUFFI G., 2017. La ponctuation entre imaginaire et sentiment linguistique. *Linx* 75, 35-56.
- SKUPIEN DEKENS C.**, 2019. Ponctuation et cohésion : ce que les premiers mots nous disent. Étude sur un corpus de sermons protestants du XVI<sup>e</sup> au XVIII<sup>e</sup> siècle. *Verbum* XLI, 209-227.
- VANDENDORPE Ch., 1999. *Du papyrus à l'hypertexte. Essai sur les mutations du texte et de la lecture*. Paris, éd. La découverte.
- VIBERT A., 2013. Faire place au sujet lecteur en classe [...]. <https://eduscol.education.fr/document/5687/download>
- VIRBEL J., 1989. The Contribution of linguistic knowledge to the interpretation of Text Structure. *Structured Documents*. Cambridge: CUP, 161-181.
- VITZ E. B., 1999. *Orality and Performance in Early French Romance*, Cambridge, Brewer.
- VOLPILHAC A., 2015. « *Le secret de bien lire* ». *Lecture et herméneutique de soi en France au XVII<sup>e</sup> siècle*. Paris, Champion.
- VORSTIUS C., RADACH R., & LONIGAN C. J., 2014. Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition*, 22(3-4), 458-485.
- WALTER R.**, 2019. Transcrire dans / pour un projet éditorial : le module Transcript sur la plateforme EMAN. *Atelier Digit\_Hum 2019 : Les humanités numériques en langues*, Oct 2017, Paris, France. ([hal-02363061](https://hal-02363061))
- 2019. EMAN, platform for the edition of manuscripts and digital archives. *Digital Vernon Lee*, The Open University's History of Books & Reading (HOBAR) collaboration & the International Vernon Lee Society (IVLS), Oct 2019, Londres, United Kingdom. ([hal-02359635](https://hal-02359635))
- WAQUET, F., 2003. *Parler comme un livre. L'oralité et le savoir (XVI<sup>e</sup>-XX<sup>e</sup> siècle)*. Paris, Albin Michel.
- ZUMTHOR P., 1987. *La lettre et la voix. De la « littérature » médiévale*. Paris, Seuil.