

## RAPPORT D'ÉVALUATION DE L'UNITÉ

ATILF - Analyse et Traitement Informatique de la  
Langue Française

### SOUS TUTELLE DES ÉTABLISSEMENTS ET ORGANISMES :

Université de Lorraine

Centre national de la recherche scientifique -  
CNRS

---

**CAMPAGNE D'ÉVALUATION 2022-2023**  
VAGUE C

Rapport publié le 01/02/2023



Au nom du comité d'experts<sup>1</sup> :

Gabriel Bergounioux, Président du comité

Pour le Hcéres<sup>2</sup> :

Thierry Coulhon, Président

En vertu du décret n° 2021-1536 du 29 novembre 2021 :

1 Les rapports d'évaluation "sont signés par le président du comité". (Article 11, alinéa 2) ;

2 Le président du Hcéres "contresigne les rapports d'évaluation établis par les comités d'experts et signés par leur président." (Article 8, alinéa 5).

Ce rapport est le résultat de l'évaluation du comité d'experts dont la composition est précisée ci-dessous. Les appréciations qu'il contient sont l'expression de la délibération indépendante et collégiale de ce comité. Les données chiffrées de ce rapport sont les données certifiées exactes extraites des fichiers déposés par la tutelle au nom de l'unité.

## MEMBRES DU COMITÉ D'EXPERTS

|                    |  |
|--------------------|--|
| <b>Président :</b> | M. Gabriel Bergounioux, Université d'Orléans   |
| <b>Experts :</b>   | Mme Angélique Amelot, Sorbonne-Nouvelle, Paris<br>M. Jean-Pierre Chevrot, Université-Grenoble -Alpes<br>Mme Sascha Diwersy, Université-Montpellier 3<br>M. Ludovic Tanguy, Université Toulouse 2<br>M. Johan Van Der Auwera, Université d'Anvers, Belgique |

## REPRÉSENTANT DU HCÉRES

M. Bernard Laks

## CARACTÉRISATION DE L'UNITÉ

- Nom : Analyse et traitement informatique de la langue française
- Acronyme : ATILF
- Label et numéro : UMR 7118
- Composition de l'équipe de direction : M. Alex Boulton

## PANELS SCIENTIFIQUES DE L'UNITÉ

SHS Sciences humaines et sociales  
SHS4 L'esprit humain et sa complexité

## THÉMATIQUES DE L'UNITÉ

Linguistique, lexicque, discours, didactique et sociolinguistique.

## HISTORIQUE ET LOCALISATION GÉOGRAPHIQUE DE L'UNITÉ

L'unité est une UMR placée sous la tutelle du CNRS et de l'Université de Lorraine. Elle est localisée dans un bâtiment CNRS et un autre bâtiment de l'université de Lorraine sur le Campus Lettres et Sciences Humaines.

## ENVIRONNEMENT DE RECHERCHE DE L'UNITÉ

L'unité a bénéficié des synergies sur site, en particulier des collaborations nouées avec le LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications) et l'INIST (Institut de l'Information Scientifique et Technique). Sa présence dans une UFR de lettres et langues lui a permis d'accueillir des enseignants qui sont venus renforcer le potentiel du laboratoire et complété l'offre en FLE par une ouverture à l'apprentissage des langues. Cet apport se retrouve dans les formations doctorales.

UMR dans une région qui en compte peu en SHS, l'ATILF a été un élément moteur dans le lancement d'un Equipex dont l'interruption se répercute sur le plan des ressources. L'unité n'a pas développé de lien fort avec la MSH Lorraine dont la situation n'était pas la plus favorable à un engagement.

C'est surtout sur un plan régional que la présence de l'ATILF est lisible à travers une politique de contrats qui se poursuivra dans les années à venir avec un volet du CPER. Les relations avec les établissements d'enseignement supérieur proches n'ont pas été spécifiées.

## EFFECTIFS DE L'UNITÉ : en personnes physiques au 31/12/2021

|   |           |
|---|-----------|
| <b>Personnels permanents en activité</b>                                    |           |
| Professeurs et assimilés  | 13        |
| Maîtres de conférences et assimilés   | 31        |
| Directeurs de recherche et assimilés  | 1         |
| Chargés de recherche et assimilés   | 2         |
| Chercheurs des EPIC et autres organismes, fondations ou entreprises privées | 0         |
| Personnels d'appui à la recherche   | 26        |
| <b>Sous-total personnels permanents en activité</b>                         | <b>73</b> |
| Enseignants-chercheurs et chercheurs non permanents et assimilés            | 38        |
| Personnels d'appui à la recherche non permanents                            | 6         |
| Post-doctorants   | 0         |
| Doctorants  | 49        |

|   |            |
|---|------------|
| <b>Sous-total personnels non permanents en activité</b> | <b>93</b>  |
| <b>Total personnels</b>                                 | <b>166</b> |

RÉPARTITION DES PERMANENTS DE L'UNITÉ PAR EMPLOYEUR : EN PERSONNES PHYSIQUES AU 31/12/2021. LES EMPLOYEURS NON TUTELLES SONT REGROUPÉS SOUS L'INTITULE « AUTRES ».

| Employeur              | EC        | C        | PAR       |
|------------------------|-----------|----------|-----------|
| Université de Lorraine | 44        | 0        | 3         |
| CNRS                   | 0         | 3        | 22        |
| Autres                 | 0         | 0        | 1         |
| <b>Total</b>           | <b>44</b> | <b>3</b> | <b>26</b> |

## BUDGET DE L'UNITÉ

|   |                  |
|---|------------------|
| Budget récurrent hors masse salariale alloué par les établissements de rattachement (tutelles) (total sur 6 ans)  | 1 066 316        |
| Ressources propres obtenues sur appels à projets régionaux (total sur 6 ans des sommes obtenues sur AAP idex, i-site, CPER, collectivités territoriales, etc.)  | 3 698 861        |
| Ressources propres obtenues sur appels à projets nationaux (total sur 6 ans des sommes obtenues sur AAP ONR, PIA, ANR, FRM, INCa, etc.)   | 4 232 175        |
| Ressources propres obtenues sur appels à projets internationaux (total sur 6 ans des sommes obtenues)   | 195 343          |
| Ressources issues de la valorisation, du transfert et de la collaboration industrielle (total sur 6 ans des sommes obtenues grâce à des contrats, des brevets, des activités de service, des prestations, etc.) | 392 095          |
| <b>Total (€)</b>  | <b>9 584 790</b> |

## AVIS GLOBAL

L'ATILF est une unité dont l'identité dans le paysage de la recherche linguistique en France est bien identifiée. Héritière de l'INaLF (Institut National de la Langue Française), elle a, au sein du CNRS, poursuivi les missions qui lui ont été confiées en matière de ressources textuelles (Frantext) et de dictionnaires (TLF, FEW). Ces orientations ont été prolongées par :

- (i) l'extension des ressources dictionnaires (*Dictionnaire du Moyen Français, Dictionnaire Étymologique Roman*),
- (ii) l'analyse des données, en particulier pour l'étude de la morphologie (partie Démonette de l'ANR Démonext), syntaxique (ANR PARSEME) et sémantique (programme IUF),
- (iii) la collecte de corpus avec la mise en place du Centre National de Ressources Textuelles et Littéraires (CNRTL) et, pendant quelques années, l'Equipex Ortolang,
- (iv) le traitement automatique.

L'intégration de deux unités et l'accueil de nouveaux collègues ont réorienté une partie des travaux vers les questions de didactique correspondant au domaine d'étude du CRAPEL (Centre de Recherches et d'Applications Pédagogiques en Langues) à partir de 2006, complétant l'apport du LANDISCO (Langues, Discours, Cognition) constitutif de l'ATILF dès sa fondation en 2001.

Les évolutions en cours reflètent un certain renouvellement thématique : une approche fondée sur les corpus, l'ouverture aux langues vivantes, un intérêt pour les disciplines médicales – en particulier l'orthophonie – et le poids croissant de la didactique. D'autres domaines apparaissent comparativement en retrait, notamment l'édition de textes anciens (voir néanmoins l'édition de *Cité de Dieu*), tout comme la dialectologie ou la phonétique. De même, ni le travail accompli dans ORTOLANG et le TLFi n'ont été mis en avant dans le portfolio.

L'unité a su remplir pleinement son rôle d'animation dans la recherche en région en répondant à des appels d'offre. Elle copilote, dans le cadre du Contrat de Plan État-Région, le programme « Langues, Connaissances et Humanités Numériques » (LCHN). Le partage de la direction avec le LORIA resserre des liens qui se sont distendus comme le montre le retrait de cette UMR dans la poursuite de l'Equipex Ortolang.

L'unité doit faire face à la baisse des crédits récurrents issus du CNRS qui n'ont pas été compensés par la tutelle universitaire. Les contrats obtenus ne suffisent pas à ralentir une attrition continue des ressources. L'ATILF bénéficie de la présence d'un nombre significatif de personnels d'appui à la recherche dont la contribution aux travaux de l'unité devra être précisée. Le recroisement des affectations dans les équipes et dans les services communs de l'unité ne permet pas une identification des missions en correspondance avec chacun des BiATSS, un dévers apparaissant entre l'effectif global et les affectations déclarées. Si le nombre d'enseignants-chercheurs s'est légèrement accru, celui des BiATSS est resté constant et celui des IT s'est rétracté. Dès lors que l'ATILF est investie de missions quasi-régaliennes en matière de dictionnaire et de corpus de textes de référence, une réflexion est à conduire sur ce point avec l'ensemble des partenaires. Par ailleurs, aucun CR ou DR n'a été recruté durant le contrat en cours.

L'ATILF a décidé de se structurer en cinq équipes avec quatre services d'appui (administration générale, soutien technique à la recherche, communication et valorisation, centre de documentation). Les équipes ont un potentiel d'enseignants-chercheurs très variable, de vingt-cinq à moins de trois ETP. Trois d'entre elles apparaissent comme les héritières de l'INaLF, une du CRAPEL et la dernière de LANDISCO.

L'équipe « Lexique » est au centre des pôles d'intérêt depuis la fondation de l'unité. Bien connue par le *Trésor de la Langue Française informatisé (TLFi)*, c'est aujourd'hui autour de la morphologie et des systèmes lexicaux qu'une contribution significative est apportée, selon des orientations théoriques dont la continuité dépendra des recrutements à venir. Deux revues de qualité sont éditées : *Verbum* et les *Cahiers de lexicologie*.

L'équipe « Linguistique historique française et romane » s'est consacrée principalement au *Dictionnaire du Moyen Français (DMF)* et au *Dictionnaire Étymologique Roman (DÉRom)*, des entreprises de long terme dont il n'existe ailleurs aucun équivalent, avec une certaine distension du lien à la philologie (édition de textes anciens).

L'équipe « Discours » agrège des travaux en acquisition du langage et en grammaire. L'*Encyclopédie Grammaticale du Français* (en ligne) renouvelle les méthodes d'écriture et de diffusion d'une grammaire qui se propose d'être une référence dans son domaine. La cohérence avec les études psycholinguistiques ne ressort pas clairement.

L'équipe « Didactique des langues et sociolinguistique », la plus nombreuse (près de la moitié des EC et un tiers des doctorants), fait preuve d'un dynamisme certain avec une ouverture vers une approche par corpus (FLEURON) et une réflexion sur l'intégration sociale par l'acquisition de la langue (Portail FL2I). Son poids au sein de l'ATILF modifie visiblement la pondération des investissements scientifiques de l'unité.

L'équipe « Ressource : normalisation, annotation et exploitation » assure la poursuite de Frantext et contribue au programme PARSEME. Subdivisée en trois axes avec 3 EC, 2,5 IT et 1 doctorant, elle s'avère limitée en capacité.

# ÉVALUATION DÉTAILLÉE DE L'UNITÉ

## A - PRISE EN COMPTE DES RECOMMANDATIONS DU PRÉCÉDENT RAPPORT

Le DAE détaille de façon précise la façon dont l'unité a pris en compte les recommandations formulées dans le précédent rapport. Il fournit, pour chacun des points, des précisions et des explications qui portent le témoignage des améliorations apportées. Ces éléments, qui ont été brièvement repris lors de l'exposé général en début de visite, permettent d'apprécier les efforts accomplis et parfois à poursuivre comme l'indiquent les commentaires dans les différentes rubriques *infra*.

## B - DOMAINES D'ÉVALUATION

### DOMAINE 1 : PROFIL, RESSOURCES ET ORGANISATION DE L'UNITÉ

#### Appréciation sur les ressources de l'unité

Le laboratoire a dû faire face à une diminution importante de ses dotations pérennes, notamment en ce qui concerne celles en provenance du CNRS. Il a su relever le défi grâce à une implication soutenue dans des projets financés, dont l'acquisition et la mise en œuvre entraînent certaines lourdeurs d'ordre « managérial » qui risquent d'entraver le dynamisme dont font preuve les activités du laboratoire d'une manière générale. Une stabilisation des ressources récurrentes serait justifiée et souhaitable, en particulier pour la poursuite des missions historiques de l'unité.

#### Appréciation sur les objectifs scientifiques de l'unité

L'UMR a poursuivi une politique d'ouverture qui se manifeste tant sur le plan institutionnel que disciplinaire. Elle est très impliquée dans le réseau des instances territoriales et universitaires au niveau local, de même qu'elle peut se prévaloir d'un nombre important de collaborations à l'échelle nationale et internationale. D'un point de vue scientifique, le laboratoire a opté pour une diversification de ses activités de recherche qui consiste à faire correspondre ses grands acquis historiques dans le domaine des ressources dictionnaires et textuelles avec des axes thématiques plus récents relevant de la psycholinguistique, de la linguistique clinique et de la didactique.

#### Appréciation sur le fonctionnement de l'unité

L'ATILF se caractérise par l'importance du ratio entre personnels d'appui à la recherche et chercheurs et enseignants-chercheurs, ce qui implique, avec la logistique immobilière et informatique, une mobilisation importante des ressources. Tout en conservant l'empreinte des unités qui l'ont constituée, l'unité est parvenue à fédérer l'ensemble de ses personnels et la commémoration des vingt ans de l'ATILF a permis de rendre visible les modalités d'une présence des sciences du langage en Lorraine. Les principes de l'association semblent hétérogènes et sont peu explicites.

*1/ L'unité possède des ressources adaptées à son profil d'activités et à son environnement de recherche.*

#### Points forts et possibilités liées au contexte

Avec un soutien financier évoluant dans une fourchette de 134 150 à 115 500 € pour une moyenne annuelle de 122 918 €, le contrat quinquennal qui fait l'objet de cette évaluation a vu se confirmer la baisse des dotations du CNRS initiée en 2014. Les moyens provenant de l'Université de Lorraine, la deuxième tutelle du laboratoire, sont restés, quant à eux, plus ou moins constants (moyenne annuelle : 54 801 €, comprise dans une fourchette de 53 874 à 55 400 €). L'évolution des fonds propres (moyenne annuelle : 62 514 €, comprise dans une fourchette de 10 375 à 107 996 €) a connu des fluctuations assez importantes, dues en grande partie aux problèmes

juridiques concernant la mise à disposition de la base textuelle Frantext, principale source des fonds en question. Le laboratoire a su relever ces défis en maintenant une implication soutenue dans des projets financés (moyenne annuelle : 1 357 231 €, comprise dans une fourchette de 880 473 à 1 640 943 €) relevant de différentes instances sur le plan régional, national ou encore européen. Ainsi, des chercheurs du laboratoire sont ou étaient engagés en tant que (co-)porteurs ou partenaires dans des grandes initiatives comme le CPER (Contrat de plan État / Région Lorraine) « Langues, Connaissances et Humanités Numériques » (LCHN), LUE (Lorraine Université d'Excellence), IMPACT « Open Language and Knowledge for citizens » (OLKi) ainsi que dans divers projets ANR (Derivational Morphology in Extension – DEMONEXT ; Syntactic parsing and multiword expressions in French – PARSEME-FR). La manière dont le laboratoire engage ses dépenses n'appelle pas de remarque particulière.

### Points faibles et risques liés au contexte

La tendance à la baisse des moyens pérennes (que ce soit au niveau des dotations annuelles du CNRS ou du recrutement de personnels BIATTS et CR/DR) risque de fragiliser le dynamisme dont fait preuve le laboratoire dans ses domaines de spécialité qui, avec le développement et la diffusion de ressources textuelles et dictionnaires d'ampleur, nécessitent des investissements durables en termes de main-d'œuvre et de matériel. Même si le laboratoire réussit à combler (au moins partiellement) ce manque, les lourdeurs d'ordre gestionnaire liées à l'acquisition et la mise en place de projets financés, en particulier au niveau européen, constituent une entrave non négligeable. Un autre facteur de risque susceptible d'impacter la stabilité des recettes du laboratoire concerne la situation juridique précaire de la base Frantext, qui n'a pas encore pu être clarifiée malgré des efforts poursuivis depuis plusieurs années.

## *2/ L'unité s'est assigné des objectifs scientifiques, y compris dans la dimension prospective de sa politique.*

### Points forts et possibilités liées au contexte

Poursuivant sa stratégie adoptée lors du précédent contrat quinquennal, l'ATILF a mené une politique d'ouverture tant sur le plan institutionnel que disciplinaire. Le document d'auto-évaluation (DAE) fait état d'une forte implantation du laboratoire dans le réseau des instances territoriales (Métropole du Grand Nancy, Région Grand-Est) et universitaires au niveau local (Pôle « Connaissance, Langage, Communication, Sociétés » de l'Université de Lorraine, Ecole doctorale 78 « Sociétés, Langages, Temps, Connaissances », Lorraine Université d'Excellence, Maison des Sciences de l'Homme Lorraine et le Comité de Coordination et d'Orientation Scientifique Lorrain). Pour une partie des projets financés (CPER LCHN et LUE OLKi, cf. *supra*), l'ATILF a été partenaire ou co-porteur avec le laboratoire INRIA LORIA, autre acteur, sur le site nancéen, dans le domaine du traitement de données langagières. Il est également impliqué dans un grand nombre de collaborations avec des organismes de recherche à l'échelle nationale (les laboratoires LLF, CLLE, STL, etc.) et internationale (Allemagne, Belgique, Suisse, Portugal ; Canada, Brésil ; Chine, Koweït). Sur le plan scientifique, la politique du laboratoire a consisté à inscrire ses activités dans les priorités thématiques de ses tutelles (concernant notamment les Humanités numériques et les « big data » d'une part, et les défis sociétaux liés aux troubles du langage et l'apprentissage des langues, de l'autre), ce qui a induit une diversification disciplinaire assez prononcée dont l'objectif est d'aligner les acquis historiques de l'ATILF dans le domaine des ressources textuelles et dictionnaires avec des travaux relevant de la psycholinguistique, de la didactique et de la linguistique clinique. La structuration du laboratoire et son organisation reflètent explicitement ces choix stratégiques. Des efforts sont réalisés pour rapprocher les dimensions TAL et corpus et certains travaux de l'équipe "Didactique des langues et Sociolinguistique" (par exemple la constitution de ressources numériques multimédias pour l'enseignement du FLE). Ce processus d'intégration est cohérent avec la nouvelle structuration en thématiques croisées projetée par la future direction.

### Points faibles et risques liés au contexte

Toute louable que soit la volonté d'assumer une diversification scientifique permettant d'aller au-delà des thématiques « anciennes », la mise en œuvre d'une politique d'ouverture pluri-, voire interdisciplinaire, revêt le risque d'une forte dispersion susceptible de brouiller le profil global de l'unité. À en juger par différents propos du DAE, le laboratoire est bien conscient de ce risque, qu'il tente de contrer en mettant en place des méthodologies censées intégrer des approches exploratoires (basées sur l'exploitation, notamment quantitative, de corpus textuels) et confirmatoires (basées sur le traitement statistique de données expérimentales). D'un point de vue institutionnel, l'ATILF fait face aux conséquences liées à la fin du financement PIA de l'Equipex ayant permis la réalisation de la plate-forme ORTOLANG, qu'il s'agira de stabiliser afin d'assurer son rôle d'équipement comme facteur structurant du « paysage » de la gestion et de la diffusion de ressources langagières au niveau national. À cet égard, le laboratoire devrait clarifier le positionnement de la plate-forme face à la TGIR Huma-Num, à COCOON et aux infrastructures européennes CLARIN et DARIAH et se tourner plus résolument vers le domaine international.



### *3/ Le fonctionnement de l'unité est conforme aux réglementations en matière de gestion des ressources humaines, de sécurité, d'environnement et de protection du patrimoine scientifique.*

#### Points forts et possibilités liées au contexte

Parmi les 76 permanents du laboratoire au 31/12/21, le rapport femmes/hommes est de 63,2 %. Cela représente 48 % des ITA/BIATSS (12/13) et 70,6 % des E/C/EC (36/15).

Le laboratoire est doté d'un Conseil de Laboratoire, composé de 15 personnes. Le DU le préside et les 2 directeurs-adjoints sont membres de droit. L'élection des 10 membres se fait au sein de collèges représentatifs du personnel. Le collège ITA/BIATSS a 4 sièges. Le fonctionnement de l'ensemble n'appelle pas de remarques particulières.

Le personnel ITA/BIATSS est en charge des missions spécifiques liées aux réglementations et en conformité avec celles-ci : ont été désignés un assistant de prévention, un correspondant bâtiment, un correspondant SSI, un correspondant formation et un correspondant pour l'égalité professionnelle et la parité.

La Commission Locale d'Hygiène, de Sécurité et de Conditions de Travail est présidée par l'agent de prévention. Y figurent plusieurs ITA/BIATSS. Un IE est en charge des questions de développement durable et de l'écologie. En plus du recyclage et de la réutilisation des déchets, du papier ou des consommables (cartouches, piles), il est impliqué dans la politique de missions et le renouvellement des matériels.

L'UMR témoigne d'une attention à la situation de ses personnels et aux échanges avec ses partenaires en prenant en compte les conditions de travail et les modalités de collaboration. Une réflexion et des concertations intègrent les attentes de l'écologie de la recherche et favorisent l'élaboration de bonnes pratiques en matière de publications et de données ouvertes ou la création de plateformes pour le partage de ressources.

Une mutualisation interne des moyens immobiliers, informatiques et humains est recherchée et l'attention se porte sur les conditions de travail (cf. un dossier QVT (Qualité de Vie au Travail) validé par le CNRS au printemps 2021) comme en témoignent l'acquisition de fauteuils ergonomiques, des accès facilités pour les personnes à mobilité réduite, etc.).

L'unité s'est montrée vigilante pendant la crise Covid-19, attentive aux risques psycho-sociaux, tant pour les doctorants et personnels non permanents que pour les titulaires. Ce souci se retrouve dans l'encadrement et le suivi des doctorants. L'assistant ingénieur, assistant de prévention, a été particulièrement sollicité pendant la Covid-19. Il a veillé aux conditions matérielles de télétravail et a préparé le « Plan de Reprise d'Activité » (PRA) à la sortie du premier confinement et le plan de continuité de travail (PCA).

Les doctorants ont confirmé la vigilance de l'unité concernant leurs conditions de travail (mise à disposition d'un bureau) et le suivi (comité de suivi de thèse).

#### Points faibles et risques liés au contexte

Le rapport indique que les données utilisées au laboratoire nécessitent une attention particulière en ce qui concerne la sécurité informatique, gérée par le service STR et par le Correspondant Sécurité des Systèmes Informatiques (CSSI). Des indications supplémentaires sur la protection des données permettraient d'apprécier la robustesse des solutions retenues. Enfin, le rapport ne détaille pas les procédures suivies pour une mise en conformité avec les exigences de la RGPD, ni d'éventuels contacts avec la CNIL.

En dépit d'un ratio Hommes/Femmes exemplaire, on relève, chez les ITA/BIATSS, un léger déséquilibre dans la répartition entre les corps : les femmes représentent 60 % des techniciens (6/4), mais 40 % des IR, IE et AI (6/9).

La description de la gestion du télétravail après la période Covid-19 n'est pas faite dans le rapport et c'est dans la présentation lors du comité de visite qu'elle a été explicitée. Selon la réponse apportée par le conseil de laboratoire, l'évolution des relations entre les agents et les affectations dans l'exercice des fonctions pourrait être profondément transformée, un point qui mérite qu'on y prête attention.

Une difficulté dans le recrutement tient à l'adéquation de certains profils avec les exigences de la recherche : les limites de recrutement du master sciences du langage ne favorisent pas des recrutements en TAL et les domaines d'intervention des ITA ménagent des lacunes dans la couverture de nouveaux besoins comme le traitement statistique des données.

## DOMAINE 2 : ATTRACTIVITÉ

### Appréciation sur l'attractivité

L'ATILF est l'un des piliers de la recherche linguistique en France, implantée dans sa région avec un fort potentiel à l'échelle nationale et internationale. Sa présence dans le domaine des corpus écrits (Frantext, CNRTL) et de la lexicographie (TLF) s'est accompagnée d'un travail théorique en morphologie (Démonette), lexicale (théorie Sens Texte, ANR PARSEME) et syntaxe (EGF). L'ouverture à la didactique (FLEURON), à la psycholinguistique et à la clinique, à la traductologie a élargi le champ d'intervention. Les évolutions démographiques, financières et thématiques appellent réflexion.

*1/ L'unité est attractive par son rayonnement scientifique et contribue à la construction de l'espace européen de la recherche.*

#### Points forts et possibilités liées au contexte

Les projets phares retenus dans la présentation des réalisations par équipes sont représentatifs des orientations de la recherche telles que l'unité a choisi de les affirmer. L'ATILF a su maintenir les chantiers majeurs qui ont assuré sa réputation en matière de philologie, qu'il s'agisse de bases de données :

- actualisation de Frantext et refonte du système d'exploitation,
- construction de Démonette dans le cadre de l'ANR Démonext,
- exploitation des corpus pour l'enseignement du FLE (FLEURON),

de dictionnaires :

- *Dictionnaire du Moyen Français* (DMF),
- *Dictionnaire Étymologique Roman* (DÉRom)

ou d'études sur le lexique :

- système de graphes dans la perspective de la théorie Sens / Texte,
- expressions polylexicales (PARSEME),

et la grammaire

- Encyclopédie Grammaticale du Français (EGF).

De nouveaux projets ont pris le relais qui bonifient les compétences de l'unité en se consacrant à des champs d'investigation en littérature moderne, en orthophonie (Orthocorpus) ou au discours pathologique. Deux points saillants ressortent de cet inventaire dans son ouverture : d'une part, la capacité à conserver à l'ATILF son rôle éminent en matière de lexicologie et de lexicographie, tant dans la réalisation de produits que dans le retour sur ses propres pratiques. La reprise des *Cahiers de lexicologie* en témoigne. D'autre part, une créativité dans des réalisations qui ménagent un espace propice à l'accueil de nouveaux collègues.

La reconnaissance à l'international des expertises spécifiques est manifeste dans la présence de doctorants et plus encore de chercheurs internationaux : une trentaine ont été accueillis, pour des séjours de durée variable, pendant les années du contrat. La liste des pays témoigne de la variété des échanges. Réciproquement, on comptabilise environ quatre-vingt-dix conférences à l'étranger durant cette période. Cette implication, relayée par un Master ERASMUS, se retrouve dans la négociation de programmes de recherche impliquant principalement la Suisse, l'Espagne et l'Allemagne. Cette politique d'ouverture bénéficie d'un soutien significatif de la région, en particulier dans le cadre du Contrat de Plan État-Région où les résultats obtenus constituent l'indice certain d'une reconnaissance régionale attestée par des abondements issus du Grand Nancy et de la Lorraine.

Bien ancré dans son espace géographique, pourvu de personnels compétents en appui à la recherche par le CNRS, bien positionné sur des spécialisations clairement identifiées avec des ouvertures stimulantes, l'ATILF s'est manifestement donné les moyens de poursuivre les missions essentielles fondatrices de son unité et d'assurer la pérennité de ses travaux.

#### Points faibles et risques liés au contexte

L'ambition initiale du TLF de concevoir un dictionnaire de référence de la langue française a été amplifiée par une mise à disposition du contenu sur Internet qui n'avait pu être anticipée au moment de son lancement. Au changement technologique fait écho le défi social d'une consignation des variétés du français, amorcée grâce à une collaboration avec le Québec sans qu'il ne soit fait mention d'une extension aux multiples formes prises par la langue à l'échelle mondiale. Alors que le relevé des items reste contraint par la priorité donnée à l'écrit, la réduction de l'empan géographique appelle une réflexion sur les références nécessaires à un dictionnaire du français vivant. Qu'il s'agisse des usages hors métropole (y compris ultramarins) – on pense au

travail accompli par le *Dictionnaire Des Francophones* avec lequel le partenariat peut être prolongé au-delà de l'inclusion du *Dictionnaire des régionalismes de France* – ou des emplois non standard, des ressources importantes de l'unité en matière de corpus oraux et de sociolinguistique n'ont pas été intégrées à la hauteur des attentes en la matière.

Un autre « monument » est absent des quinze programmes mis en avant : la continuation du *FEW*, ce que confirme la lecture du document d'auto-évaluation qui n'accorde pas une place significative à cette entreprise lexicographique. Grâce à des subsides de la Suisse, le projet continue sur sa lancée et contribue à l'enrichissement d'autres actions conduites par l'unité, sans que soit observable une dynamique particulière autour de ce programme. La relation avec la partie étymologique du *TLFi* et plus encore avec le *DÉRom* reste imprécise.

La diversité des recherches, inhérente à une unité de recherche dont l'université de rattachement entend qu'elle regroupe un maximum de ses personnels d'enseignement, fait courir le risque d'une certaine dispersion qui dilue les objectifs poursuivis collectivement entre la littérature moderne – la philologie des textes plus anciens est un héritage nécessaire de l'histoire de l'ATILF – et les domaines de la santé (entre autres l'orthophonie). L'importance croissante de la didactique, particulièrement visible en matière de publications, si elle est un relais important dans le domaine des applications, réoriente une partie des activités dans des directions qui s'avèrent moins accordées aux intentions premières. Une plate-forme comme *EDOlang* est en continuité des travaux conduits sur l'anglais et l'allemand et les travaux conduits sur le FLE reflètent l'intégration du *CRAPEL* au risque d'un éloignement des études qui sont les spécificités de l'ATILF dans le paysage des sciences du langage.

Les relations internationales, quoique diversifiées par pays et par domaine, ne font pas apparaître de programmes bilatéraux (ou multilatéraux) qui dessineraient les contours d'un réseau structuré à même de porter à terme des programmes européens d'ampleur. La même remarque pourrait être faite concernant la France. Qu'il s'agisse du *CNRTL* ou de l'*Equipex Ortolang*, la préservation à long terme du potentiel acquis semble fragile. Ainsi, les crédits du *PIA* n'étant plus renouvelés, l'*Equipex* est mentionné à présent comme un programme de l'ATILF sans qu'ait été réellement poursuivie la collaboration avec l'ensemble des unités fondatrices, en particulier le *LORIA*.

## 2/ L'unité est attractive par la qualité de sa politique d'accueil des personnels.

### Points forts et possibilités liées au contexte

Deux éléments étaient à prendre en compte durant le contrat, l'un concernant les mouvements de personnels, enseignants-chercheurs, IT et BiATSS, l'autre résultant de la fusion des universités de Nancy et de Metz. La lecture du rapport tend à montrer qu'il ne s'est pas présenté de difficulté particulière dans ce domaine et, à en juger par les tableaux fournis au Hcéres, alors que l'on compte quarante-sept chercheurs et enseignants-chercheurs à la date de rédaction du rapport (les chiffres sont repris de ceux figurant à la page 7 du document d'auto-évaluation), et sans comptabiliser les départs en retraite, deux seulement ont effectué un changement d'unité avant leur retraite, signe fort d'un attachement au fonctionnement collectif. Qu'il faille y voir l'effet d'une attractivité de l'ATILF auprès de MCF en poste dans d'autres disciplines ou des encouragements de l'établissement à accroître la taille de ses composantes, le départ de MCF a été compensé, et au-delà, par l'arrivée de neuf collègues de même rang, le nombre de professeurs restant constant, soit une augmentation de 6 % dans un environnement partout contraint par des budgets en rétraction. À cela s'ajoute la présence de six émérites (quatre précédemment) dont 1 CNRS et 2 postes d'ATER. En l'absence d'indications chiffrées dans le dossier sur l'évolution des effectifs étudiants dans les filières concernées et l'élargissement des champs d'intervention, il est difficile d'apprécier quelles sont les conséquences en matière de taux d'encadrement, un indicateur précieux pour évaluer la disponibilité réelle en recherche.

Concernant les personnels d'appui à la recherche, la réduction concerne deux postes d'ATR. Les personnels CNRS et les 3 BiATSS de l'Université de Lorraine permettent à l'ATILF de bénéficier de compétences importantes en interne, avec des ingénieurs et assistants-ingénieurs rompus au travail sur les types de programmes développés par l'unité et à même d'accompagner les engagements des contrats. Le personnel CNRS comprend 3 Assistants Ingénieurs, 3 Techniciens, 7 ingénieurs d'Étude, mais seulement 3 ingénieurs de recherche dont 2 à 50 %.

Le personnel CNRS et BiATSS est impliqué dans diverses missions spécifiques : assistant de prévention, correspondant bâtiment, correspondante égalité, diversité, inclusion, correspondante égalité pour la Mission pour la place des femmes au CNRS (COREGAL) de la Délégation CNRS Centre-Est, correspondant de la sécurité des systèmes d'information, correspondante valorisation INSHS CNRS. Plusieurs personnels ont été promus entre 2016 et 2021 notamment par concours interne : IRHC, IR2, IE HC, IE, AI, TCE, TCN et ATP1. Les ITA/BiATSS sont très impliqués dans les séminaires du laboratoire.

L'attractivité se retrouve dans l'augmentation régulière du nombre de doctorants (et du nombre de thèses) ainsi que dans les sollicitations extérieures pour la participation à des jurys de thèse, à des conférences ou à des séminaires. L'inventaire constitue une preuve de l'amplitude du spectre des échanges, en France et à l'international.

### Points faibles et risques liés au contexte

Un premier risque apparaît à la lecture du rapport concernant la démographie des professeurs et des personnels d'appui à la recherche. À la fin de 2022, huit professeurs auront soixante ans et plus et, à raison d'un par an, quatre autres franchiront ce cap dans les quatre années suivantes. 4 professeurs, nés entre 1953 et 1955, mentionnés dans le fichier Excel, ont déjà dû partir lors du contrat en cours. Si le renouvellement des cadres offre des possibilités pour une nouvelle dynamique, il rend patentes les limites de la formation en interne puisque, pour les années prises en compte, seulement 4 HDR ont été soutenues dont le champ d'exercice concerne la didactique, la traductologie et la littérature. Selon le directeur, deux autres sont proches et quatre semblent possibles dans les années à venir. Si des IR et des MCF ont apporté la preuve de leurs capacités à gérer des programmes ou à participer à l'encadrement de thèses, la question se pose concernant la continuité d'actions conduites depuis plusieurs décennies pour lesquelles il est difficile de trouver un rassemblement de force comparable à celui dont bénéficie l'unité. Les spécialisations requises dans les domaines de spécialisation historiques n'ont pas été encouragées en France et le vivier se trouve aujourd'hui plutôt en dehors des frontières.

Un même effet générationnel se retrouve du côté des personnels CNRS où des recrutements concomitants fragilisent une reconduction alors qu'aucun appui n'est venu en renfort au moment de la fin de l'Equipex Ortolang. Concernant le personnel CNRS, 6 personnes ont été recrutées ou promues entre 2016 et 2021 et deux postes n'ont pas été remis au concours. Sur les sept postes d'ingénieurs d'étude, deux sont en CDD. Entre 2016 et 2020, en plus de la perte de deux postes de soutien, on note également la perte d'un poste de contractuel (de 6 à 5 : 2 IR et 3 IE).

Comme la direction de l'unité le souligne, l'absence de recrutement de CR ou de DR, en dépit de candidatures de qualité, est d'autant plus préoccupante qu'elle s'inscrit dans une affectation des postes à un nombre restreint d'unités sur des problématiques qui, en n'accordant pas toute l'attention requise aux grands programmes de recherche pluriannuels au profit de thématiques très spécialisées, fragilisent la conduite des projets à long terme. Une unité dont les points forts concernent la langue française (et des publications correspondant à cette orientation) est handicapée dans les profils qu'elle accompagne alors que les moyens accordés en personnels d'appui appellent une ambition à la même hauteur en termes de recherche fondamentale et appliquée.

Il est de la seule responsabilité de l'unité de définir et d'organiser ses équipes. La variété des engagements et la définition d'objectifs spécifiques ont minoré l'importance du CNRTL et d'Ortolang malgré l'ambition affirmée d'être à l'échelle nationale le relais des infrastructures européennes, DARIAH et CLARIN.

### *3/ L'unité est attractive par la reconnaissance que lui confèrent ses succès à des appels à projets compétitifs.*

#### Points forts et possibilités liées au contexte

Le compte d'exploitation ne fait pas ressortir suffisamment le détail des ressources nettes de l'ATILF par source de financement et par année. Quoi qu'il en soit, l'unité s'est réellement distinguée dans l'obtention de fonds sur contrats, fléchant les crédits obtenus à la réalisation de tâches qui auraient été couvertes, en d'autres temps, par des crédits récurrents lesquels, à présent, sont réservés prioritairement à la rémunération de personnels permanents dont les missions sont confortées par le recours à des contractuels. En cumul sur les années couvertes par le contrat, pour 9,6 millions d'euros de ressources financières – hors salaires des permanents et assimilés –, le CNRS a apporté environ 8 % des ressources (un tiers en fonctionnement et deux tiers en équipement) et l'université 4 % (en fonctionnement dans son intégralité).

L'ATILF a prouvé sa capacité à remédier à cette situation par l'obtention de crédits de l'ANR en assurant le pilotage de trois projets : PARSEME, DEMONETTE et un programme ANR DFG (OEF). Les sommes indiquées dans le dossier globalisent la destination des fonds (environ deux millions d'euros au total) dont une partie est destinée à la redistribution entre les partenaires. L'implication dans un certain nombre de programmes extérieurs (dont d'autres obtenus auprès de l'ANR) contribue à l'accroissement du bilan. Auparavant, les sommes versées dans le cadre du PIA (Equipex Ortolang) avaient représenté un apport très important en contribuant à promouvoir une politique orientée vers les corpus.

La qualité de l'intégration dans l'environnement régional montre explicitement le degré de reconnaissance accordé par ces instances puisque des abondements ont été obtenus à l'échelle de la nouvelle région (Grand Est), de la Lorraine et de l'agglomération de Nancy avec un apport d'importance négocié avec succès dans

le cadre du Contrat de Plan État-Région sous l'intitulé LCHN (Langues Connaissances et Humanités Numériques). Cet apport permet au bilan financier d'amortir la dégrèvement régulière de l'abondement en crédits récurrents de la tutelle CNRS avec un complément conséquent des institutions de recherche de la Suisse (en particulier pour le FEW) et, à un moindre degré, de la Belgique.

D'autres financements ont été obtenus dans le cadre de partenariats avec l'étranger, en premier lieu avec le Québec, le Brésil et l'Espagne, sans exclusive de contrats à financements plus limités issus de l'Allemagne, du Koweït et de la Chine. Il faut ajouter à cela les contrats doctoraux, une quinzaine auprès de l'ED de rattachement, 2 CIFRE, 6 en provenance d'autres EPST et 1 de la Région, en plus des thèses financées par le pays d'origine.

### Points faibles et risques liés au contexte

La situation financière du laboratoire est saine et la gestion des fonds correspond aux critères attendus dans le domaine. Il est en revanche inquiétant de voir la chute brutale du budget, de l'ordre d'un tiers puisqu'il est passé d'une moyenne de 1,5 million d'euros pendant les quatre premières années du contrat à moins de 1,2 million d'euros en 2020 et en dessous de 1 million d'euros en 2021, une diminution peut-être en partie liée au COVID. Le déclin appelle d'autant plus à la vigilance que la politique scientifique de l'ATILF qui, sur la période concernée, aboutit à un cumul proche de dix millions d'euros, n'est soutenable qu'en raison d'une prise en charge importante de nombreuses fonctions administratives et techniques par les IT et de BiATSS qui constituent un élément essentiel de continuité des actions entreprises.

Le devenir de l'Equipex Ortolang est symptomatique à la fois d'une politique nationale qui se désintéresse des programmes impulsés sans envisager leur continuité, au risque de perdre des compétences et des réalisations, et d'un certain recentrage de l'unité qui a choisi de restreindre l'empan de son réseau plutôt que de chercher à l'accroître. Parallèlement, une optimisation des moyens aurait impliqué de discuter les modalités d'une collaboration avec l'instance centrale dans le dispositif actuel de la recherche en matière d'entrepôt, autant de corpus que de bases de données : la TGIR Huma-Num n'est mentionnée qu'à deux reprises dans le rapport, en particulier à travers le consortium CORLI, surtout comme une instance de dialogue.

On relève un déséquilibre entre l'engagement régional et le rôle potentiellement dévolu à l'ATILF dans le pilotage des initiatives à l'échelle nationale, romane et au-delà. Les nombreuses coopérations, effectives, relèvent plus souvent de partenariats que de réseaux, renouvelant sur des opérations à durée limitée l'obtention de ressources. Ainsi, à court et moyen terme, c'est dans un cadre essentiellement régional qu'est assuré l'équilibre budgétaire de l'unité, un constat que soulignent encore la faiblesse des crédits obtenus à l'international (3 500 € dans les quatre dernières années) ou la part des ressources commerciales, forcément réduites dans le domaine des SHS, a fortiori du fait de la mise à disposition gratuite de l'ensemble des productions (les efforts en matière de science ouverte sont de grande qualité), mais néanmoins en augmentation globale.

## *4/ L'unité est attractive par la qualité de ses équipements et de ses compétences technologiques.*

### Points forts et possibilités liées au contexte

Sur ce point, l'ATILF dispose d'un des plus forts potentiels dans la cartographie de la recherche en France, tant par la disponibilité en matériels que par la spécialisation des personnels d'appui à la recherche et par l'implication de nombreux chercheurs et enseignants. Cette orientation se retrouve dans l'intervention au sein des différents axes et jusqu'à la définition d'une équipe spécialisée en Ressources. Au-delà des questions d'infrastructure, prises en charge par le groupe du Soutien Technique à la Recherche (STR), avec une vigilance sur les questions de sécurité et le souci de répondre aux attentes des chercheurs et des doctorants, des prolongements en matière de documentation (Centre de Documentation Michel Dinet (CDD-MD)) et de valorisation sont notables.

Tirant parti d'une expertise bien établie en matière de ressources textuelles (Frantext, CNRTL) et de lexicographie (TLF), l'ATILF a exploité son parc informatique et ses capacités de mise en ligne, en bénéficiant des retombées d'Ortolang et des compétences acquises pour assurer la disponibilité de bases de données (e.g. Démonette) et de corpus (e.g. TermiTH) établis selon les critères requis, en particulier la possibilité de travail collaboratif sur des plates-formes multifonctionnelles.

Les Ingénieurs d'étude et de recherche sont très impliqués dans la programmation informatique qui a abouti à la mise à disposition de 16 logiciels open-source permettant l'interaction avec les bases de données. Ils gèrent notamment la veille informatique, les abonnements et répondent aux demandes et questions. Ils sont très présents dans la valorisation et la médiation scientifique.



Les connaissances acquises dans ces domaines ont été transférées tant en philologie (projet *Cité de Dieu*) qu'en lexicographie (*DMF, DÉROM*) démontrant la pertinence des choix effectués. Une ouverture vers l'expérimentation s'est manifestée au sein de l'équipe "Discours" en fonction d'une approche neurocognitive qui exploite des appareillages d'eye tracking et l'EEG. Dans une perspective applicative a été créé un groupe d'enseignement des langues en réalité virtuelle qui se consacre aux environnements immersifs en 3D. À travers une collaboration avec Lyon est engagée l'étude des assistants vocaux dans l'apprentissage des langues.

L'équipe spécifiquement dévolue aux « Ressources : Normalisation, Annotation et Exploitation » déclare son intérêt particulier pour la constitution, le traitement, l'annotation et l'analyse des corpus avec pour premier domaine d'activité la base Frantext et pour adossement Ortolang. La poursuite de ces actions lors du contrat à venir est annoncée dans un « axe transversal méthodologique » rebaptisé « Modélisation, ressources et traitement informatique » dont les attendus sont ceux développés en TAL. Le soutien financier apporté par la région dans ce champ d'activité est important, confirmant l'enracinement de l'unité dans son environnement immédiat.

### Points faibles et risques liés au contexte

Une interrogation cruciale concerne la préservation des moyens accordés en personnels d'appui à la recherche où, par suite de considérations démographiques, se pose la question d'un renouvellement des supports de postes. Il relève de la politique d'affectation de l'InSHS CNRS que plusieurs des programmes puissent se poursuivre dans des conditions comparables à celles qui ont décidé de leur engagement.

Au-delà de la reconnaissance globale de la qualité du travail accompli, une interrogation sur la capacité de l'unité à se déployer au-delà de son organisation interne autour de programmes impliquant des compétences et des engagements d'équipements similaires demeure. Sur site, si le LORIA est mentionné à différentes reprises, un seul programme est mentionné comme étant d'intérêt commun, l'ATILF s'appuie sur une double compétence forte en sciences du langage et en informatique alors même qu'on relève peu de coopérations avec d'autres unités de l'Université de Lorraine dont les programmes concernent la médecine ou l'esthétique. Les ambitions affichées dans le domaine du TAL ne sauraient se faire en dehors d'un partenariat et le plus évident serait le plus proche, y compris dans la participation au GDR LIFT.

Il paraît significatif qu'Ortolang soit avant tout conçu comme un entrepôt de données alors que la partie « outillage » était l'un des objectifs essentiels au moment de sa labellisation. Ne fonctionnant plus qu'avec 4 des 6 laboratoires présents à sa création, l'Equipex n'a pas redéfini sa place dans le paysage de la recherche alors que ses engagements font doublon avec les missions qu'assure aujourd'hui Huma-Num.

De même, la MSHL apparaît sporadiquement, plutôt comme une case dans l'organigramme ou un pourvoyeur secondaire de fonds qu'à titre de partenaire. La plate-forme Scripto, où il était attendu que l'ATILF joue un rôle important, n'est pas mentionnée.

## DOMAINE 3 : PRODUCTION SCIENTIFIQUE

### Appréciation sur la production scientifique de l'unité

La production de l'ATILF est diversifiée, exigeante, quantitativement satisfaisante et bien répartie entre les équipes. Elle vise à respecter les principes de la science ouverte et les doctorants y contribuent activement. Les procédures de recherche garantissent la fiabilité et la reproductibilité des résultats. L'unité bénéficie d'une renommée mondiale en matière de collecte, de traitement, de partage et d'hébergement des données textuelles à visée lexicologique, tout en s'ouvrant à d'autres thématiques. Des recrutements sont nécessaires qui lui permettront de viser le meilleur standard parmi les équipes internationales spécialisées dans le traitement quantitatif des données langagières.

### *1/ La production scientifique de l'unité satisfait à des critères de qualité.*

#### Points forts et possibilités liées au contexte

Les membres de l'unité, sous l'impulsion de l'équipe de direction, font la preuve de leur capacité à mettre collectivement en œuvre une production scientifique diversifiée et exigeante quant à sa qualité.

La diversification prend plusieurs formes. Premièrement, la diversité des langues de publication atteste de l'empan international du lectorat visé. Par exemple, 23 % des articles de revues sont écrits dans une autre langue que le français, dont les trois quarts en anglais et le reste en allemand, espagnol, portugais et russe. Cette

diversification des langues est observée pour tous les types de publications. Deuxièmement, à part égale avec la diffusion de ses résultats, l'unité poursuit une activité continue, soutenue et renouvelée de constitution et d'exploitation de bases de données et de ressources, partagées avec les membres des communautés scientifiques. Cette activité fait partie de l'histoire de l'ATILF, mais le récent « tournant des données » lui confère une place centrale dans le paysage scientifique. Troisièmement, les supports de publication sont variés. L'objectif de privilégier les articles de revues a été atteint sans compromettre l'équilibre avec les autres types de publications. Finalement, les revues elles-mêmes sont de nature diverse : revues généralistes sélectives (*Nature Scientific Report*, *Frontiers*), revues phares des domaines (*Cahiers Chronos*, *Child Development*, *Computational Linguistics ; Language, Cognition and Neuroscience ; Language Learning ; Language Teaching*), revues françaises de très bon niveau (*Les Langues Modernes ; Linx ; Verbum ; Langue Française*).

L'exigence en matière de publication résulte d'une convergence entre le dynamisme des membres et le pilotage de la direction et des instances consultatives. Les membres de l'ATILF reconnaissent la nécessité de privilégier « les articles de revues, les publications auprès d'éditeurs internationaux, des textes dans une autre langue que le français, des supports accessibles en ligne » (DAE, p. 12). Mais ils ne négligent pas les publications plus locales, valorisent la production en langue française et restent attentifs aux spécificités des domaines (par exemple, la prévalence des conférences en TAL).

### Points faibles et risques liés au contexte

ATILF est une unité dynamique qui sait se doter de ressources propres permettant l'embauche de contractuels. Certains des risques qui la menacent se situent en amont des processus de recherche et d'autofinancement. Ils résultent de la baisse, sur le long terme, du nombre de chercheurs et de personnels de soutien à la recherche, qui n'est pas compensée par la hausse du nombre d'EC affiliés à l'Université de Lorraine. Notamment et surtout se fait sentir le besoin urgent d'un *Data scientist* et d'un développeur web, qui permettraient à l'unité de se situer au meilleur niveau des institutions mondiales de recherche spécialisées dans le traitement quantitatif des données langagières.

L'idée selon laquelle 'les discours et l'argumentation [dans une langue autre que le français] sont parfois très éloignés des pratiques en France et en français' (DAE, page 12) est un obstacle à la publication dans les revues de diffusion internationale, souvent en anglais. La discussion collective sur ce point doit se poursuivre.

## 2/ La production scientifique est proportionnée au potentiel de recherche de l'unité et répartie entre ses personnels.

### Points forts et possibilités liées au contexte

Sur le plan quantitatif, l'unité a produit 1182 publications écrites, orales ou affichées (DAE, tableau 5 dernière colonne). Si on projette ce total sur le potentiel de 131 emplois équivalents temps plein (DAE, tableau 2) incluant, CR/DR, EC, BIATSS et ITA, doctorants et docteurs, associés, émérites), on aboutit à 9 productions orales ou écrites au minimum par membre pendant la période de référence, ce qui témoigne d'un fort dynamisme en matière de diffusion des résultats. Ce dynamisme est partagé par les cinq équipes, le nombre de productions par membre variant peu d'une équipe à l'autre (globalement de 8 à 11). Il s'étend également aux 17 doctorants qui ont soutenu pendant la période de référence : 76 productions sont issues de leur thèse (Annexe 1. Produits et activités de la recherche, p. 117 et sv.) soit une moyenne de 4.47 productions par nouveau docteur. Pour 79 % (60/76) de ces productions, le jeune chercheur est premier auteur ou seul auteur.

Le DAE souligne que tous les membres sont encouragés à publier quel que soit leur statut et que des moyens nécessaires et suffisants sont investis dans ce sens. L'actuelle équipe de direction a repris l'initiative de ses prédécesseurs pour favoriser la publication dans des revues dites stratégiques (bien diffusées et reconnues, mais sélectives). Une liste de revues de ce type a été établie pour chaque domaine en collaboration avec les équipes. Les résultats sont au rendez-vous puisque, depuis janvier 2016, 106 articles ont été publiés dans ces revues (soit plus de 45 % des articles produits pendant la période de référence). En outre, des aides à la publication sont attribuées en interne sur appel à projet pour financer des ouvrages ou des articles en accès ouvert accompagnés d'*author processing charges*. Enfin, une enveloppe – d'ailleurs peu utilisée – est destinée à financer traduction et relecture des papiers écrits dans une langue étrangère.

### Points faibles et risques liés au contexte

Comme le mentionne le DAE, il est attendu que les enseignants-chercheurs et chercheurs titulaires publient davantage, du fait des obligations liées à leur statut. Toutefois, les jeunes chercheurs, doctorants et post-doctorants, sont dans une nécessité similaire pour faire face aux indécisions qui pèsent sur leur avenir professionnel dans un marché de recrutement particulièrement tendu. Les 17 doctorants de l'ATILF qui ont soutenu depuis 2016 (cf. Annexe 1. Produits et activités de la recherche, p. 117 sq.) ont contribué efficacement à la diffusion des résultats de leur thèse avec en moyenne plus de 4 productions par personne. Mais cette moyenne cache des disparités individuelles : 5 d'entre eux ont 2 publications ou moins à leur actif, alors que

quatre d'entre eux en ont entre 7 et 10. Il serait donc efficace que l'unité se dote d'un dispositif de suivi et d'aide aux jeunes chercheurs, incluant doctorants, docteurs et post-doctorants, afin que tous soient soutenus pour valoriser les résultats de leur recherche. Par ailleurs, toutes les équipes ne mentionnent pas le nombre exact de doctorants que les directrices et directeurs de thèse ont accompagné jusqu'à la soutenance pendant la période de référence. L'explicitation de cette répartition permettrait de mieux apprécier la distribution de l'effort de formation à et par la recherche.

### 3/ La production scientifique de l'unité respecte les principes de l'intégrité scientifique, de l'éthique et de la science ouverte.

#### Points forts et possibilités liées au contexte

Les membres de l'ATILF pratiquent une science du langage fondée sur des données. Trois thématiques transversales aux cinq équipes fondent l'unité du laboratoire : Lexique, Corpus et ressources, Aspects dynamiques du langage. L'articulation de ces thématiques illustre la solidité du projet scientifique global, qui fonde à son tour la solidité des résultats. L'étude du lexique (synchronie, diachronie, lexicographie, théorisation, modélisation notamment basée sur l'intelligence artificielle), en lien avec la constitution raisonnée et théorisée de ressources et de bases de données, continue à tenir une place centrale dans l'unité. À partir de ces données et ressources, les « membres du laboratoire (...) peuvent décrire, quantifier, modéliser, expérimenter, tester et valider de manière scientifique les hypothèses appuyées par des exemples authentiques et référencés, et cela de manière pérenne et partageable » (DAE, p.6). L'accent sur les aspects dynamiques du langage (acquisition, troubles, didactique et apprentissage, variation et changement, interaction, discours et interprétation, etc.) introduit des méthodes complémentaires, qu'il s'agisse d'approches quantitatives et formelles (expérimentation, eye-tracking, EEG, tests psycholinguistiques, analyse phonétique) ou d'approche *in situ*, plus écologiques (recherches-action, recherches-développement, approche ethno-méthodologique, collecte et traitement de données qualitatives multimodales). Les méthodologies qui sous-tendent les résultats produits par ATILF sont donc rattachées à des paradigmes éprouvés sur le plan international, incluent une part d'innovation et servent des questionnements théoriques clairs et explicites.

Sur le plan de l'éthique des protocoles de recherche, concernant la protection des personnes et de la vie privée, l'ATILF a repris les procédures de ses tutelles ; système d'information inscrit dans la démarche RGPD, formation à l'éthique assurée par l'Université de Lorraine, charte éthique pour les thèses et leur encadrement, vérification des plagiat dans les thèses et les masters de certaines composantes par des logiciels dédiés, déclaration des conflits d'intérêt. Même si l'unité appuie la mise en place d'un comité opérationnel d'éthique et d'une charte au niveau de l'établissement (deux membres de l'unité participent à la réflexion préalable), les chercheurs ne disposent pas, actuellement, de la possibilité d'obtenir un avis éthique sur leurs protocoles de recherche (le COMETS du CNRS étant une instance de réflexion, pas un comité opérationnel). Ce manque est d'autant plus préjudiciable que la collecte et le traitement des données sont centraux dans les opérations conduites par l'unité et qu'un nombre croissant de revues et d'appels à projets exigent un « ethic statement » préalable.

S'agissant de la traçabilité, de la conservation et du partage ouvert des données, les pratiques de l'ATILF sont exemplaires. Sauf lorsque ces données impliquent des droits d'auteur – auquel cas les recettes irriguent l'ensemble des travaux du laboratoire –, l'unité met les ressources qu'elle a créées ou qu'elle maintient à disposition des communautés scientifiques et du grand public, rendant un service collectif de consultation et de dépôt. En outre, l'origine des données textuelles est documentée pour la plupart des bases. En ce qui concerne le partage ouvert des publications, l'ATILF encourage le dépôt des versions d'auteurs sur HAL (« voie verte ») en accompagnement de la notice HAL, qui est obligatoire. Elle contribue à des revues en accès ouvert (« voie dorée »). Elle mesure l'impact de cette politique en utilisant l'outil Baromètre Science Ouverte fourni par l'Université de Lorraine.

L'éthique est associée à l'équité. Les statistiques sur le genre (p.26 et sv.) montrent la même tendance que dans l'ensemble des SHS : la probabilité d'accéder à des corps professionnels plus élevés dans la hiérarchie est plus importante pour les hommes que pour les femmes, aussi bien pour les BIATSS/ITA que pour les EC et les chercheurs. Des statistiques similaires ont été calculées pour l'ensemble des positions de responsabilité au sein de l'ATILF et pour les participants aux instances consultatives. Un décompte précis de l'origine nationale de l'ensemble des membres est proposé, que ces derniers soient permanents ou non. Ce suivi montre l'attention accordée à la diversité dans l'unité, dont certains aspects font l'objet de travaux de recherche. Deux chargées de mission impulsent des actions visant à faire face aux préjugés, à l'exclusion et aux processus d'invisibilisation.

#### Points faibles et risques liés au contexte

Les membres de l'ATILF sont invités à poursuivre leur action auprès des tutelles afin de promouvoir la création d'un comité d'éthique opérationnel examinant les protocoles de recherche au niveau du site (comme l'ont déjà fait de nombreuses universités françaises). À défaut d'obtenir les validations éthiques nécessaires, l'unité risque de ne pas être à même de profiter d'opportunités de financement ou de publication.



En outre, il semble nécessaire qu'une unité dont les travaux sont fondés sur la constitution ou l'hébergement de bases de données affiche plus explicitement les dispositions prises pour rendre cette activité conforme à la RGPD, en relation avec la CNIL et le DPO de l'Université de Lorraine.

Qu'il s'agisse de l'examen éthique des protocoles de recherche ou de la protection des données personnelles, les préconisations ci-dessus sont en train de devenir des standards internationaux.

## DOMAINE 4 : INSCRIPTION DES ACTIVITÉS DE RECHERCHE DANS LA SOCIÉTÉ

### Appréciation sur l'inscription des activités de recherche de l'unité dans la société

L'ATILF est un laboratoire très bien ancré dans son environnement social et culturel. La diffusion de ses productions scientifiques (données langagières, outils associés) fait partie de son ADN et continue d'être assurée et renouvelée, tant au niveau du grand public que des acteurs spécifiques. Cela se traduit à la fois par un nombre important de partenariats et de financements institutionnels et d'actions de diffusion et de communication. On regrettera toutefois le peu de relations entretenues avec le monde de l'entreprise, dans un contexte institutionnel et économique qui devrait permettre de les développer.

### 1 / L'unité se distingue par la qualité de ses interactions non-académiques.

#### Points forts et possibilités liées au contexte

L'ATILF continue à entretenir de nombreuses relations avec son environnement non-académique, en lien avec des sujets de société importants et identifiés comme prioritaires par les différents acteurs du monde de la recherche (santé et éducation).

Ces relations se traduisent en premier lieu par l'obtention de nombreux financements institutionnels, à commencer par ceux des collectivités territoriales (région et métropole) ; certains de ces financements étant obtenus dans le cadre d'appels à projets compétitifs (cf domaine 2 point 3). Ces projets concernent des thématiques de recherche variées qui peuvent relever de la recherche fondamentale sans objectif applicatif.

Au-delà de ces sources de financement qui restent a priori destinées au financement institutionnel de la recherche, plusieurs projets sont en lien avec des partenaires sur des thématiques spécifiques.

Un ensemble de travaux concerne la santé. On y trouve plus spécifiquement des actions au long terme (initiées lors du contrat précédent) en lien avec l'orthophonie (*Mocolang-o*, *Orthocorpus*, *Orthodef*). Ces travaux autour de la constitution et la valorisation de lexiques spécialisés et de corpus de textes médicaux à destination des usagers et des praticiens impliquent des acteurs comme le centre hospitalier de Bar-le-Duc et la Fédération nationale des orthophonistes, mais également un éditeur spécialisé (*Ortho-édition*).

Un autre terrain important relève du champ de la didactique et de l'apprentissage des langues en lien avec les structures éducatives (rectorat, établissements de la maternelle à l'université). Dans le même champ général, on relève des partenariats avec les ministères de l'Intérieur de la Culture pour des projets et actions visant spécifiquement les publics adultes migrants ou en situation d'illettrisme (portail *FL2I*). L'accent est mis sur le développement et la valorisation de ressources langagières et sur les interactions langagières et l'intégration de nouvelles technologies (corpus multimodaux, assistants vocaux) avec l'ambition d'être une "UMR citoyenne".

On relèvera également une collaboration avec l'Académie française autour de la numérisation et de la mise en ligne de son dictionnaire.

Le laboratoire dispose d'une structure dédiée à la communication et la valorisation de la recherche, dotée de deux personnels de soutien (niveau assistant ingénieur et technicien). Une partie de l'activité de la cellule Com-Valo concerne le développement et le suivi des différents partenariats. Le rapport d'autoévaluation montre d'ailleurs (page 15) une bonne connaissance des structures mises en place au niveau de l'institution et des tutelles et sur lesquelles les membres de l'équipe peuvent s'appuyer. Des membres de l'équipe assurent des rôles spécifiques de correspondants avec certaines de ces instances (Direction de l'Entrepreneuriat et des Partenariats Socio-économiques de l'université de Lorraine et service valorisation de l'INSHS du CNRS).

#### Points faibles et risques liés au contexte

Le rapport ne fait que très peu état de relations avec le monde de l'entreprise.

Deux thèses CIFRE sont mentionnées (sur un total de 69 doctorats dont 37 financés) dont l'une pour laquelle aucun détail n'a pu être trouvé. On ne note également aucun contrat de collaboration ni a priori de partenaires

privés dans les projets où leur présence est possible (type ANR). Au final, il apparaît que l'essentiel des collaborations et partenariats de l'ATILF, financés ou non, se cantonne au secteur public ou parapublic. On comprend la difficulté de mettre en place de tels partenariats étant donnés les thèmes de recherche développés par le laboratoire, mais il est possible de solliciter des structures (type SATT) pour accompagner le développement d'applications. Bien que le laboratoire ait identifié les différents structures et dispositifs disponibles dans l'environnement institutionnel, mais on ne trouve pas d'indication de leur utilisation concrète par les chercheurs. Il est possible que ce manque soit dû à une méconnaissance institutionnelle des activités et compétences présentes au niveau du laboratoire. On peut aussi comprendre la réticence des membres du laboratoire ou de ses instances à limiter sa liberté de recherche à des contingences économiques ou industrielles, mais le champ des sciences du langage a pu montrer la valeur ajoutée d'un intérêt scientifique pour des données et des problématiques dont l'accès se fait par le biais d'entreprises ou d'acteurs économiques au sens large.

On peut penser que plusieurs thématiques gravitant autour des données langagières et du TAL pourraient trouver leur place dans un processus de valorisation réussie avec un partenaire industriel. Ce processus pourrait impliquer un partenaire académique plus habitué à ces pratiques, comme le LORIA.

D'autre part, les relations avec les acteurs publics non-académiques sont assez limitées au niveau thématique. Elles n'impliquent donc a priori qu'une part limitée de l'effectif du laboratoire et l'essentiel de celles qui sont listées ont été initiées lors du précédent contrat. On note également une grande disparité entre les équipes, avec une absence totale d'indication d'activités tournées vers l'extérieur pour l'équipe « Linguistique Historique Française et Romane », ce qui peut en partie se comprendre au vu des thématiques.

Malgré son nom, la cellule Com-valo semble consacrer l'essentiel de son activité à la communication (voir plus bas) au détriment de la valorisation.

Le rapport indique bien que la direction du laboratoire a conscience de la faible part de ses activités (cf. tableau page 13) et la partie prospective évoque spécifiquement ce point qu'il lui faut encourager.

## *2/ L'unité développe des produits à destination du monde socio-économique.*

### Points forts et possibilités liées au contexte

L'ATILF a une longue histoire et une notoriété incontestable en ce qui concerne la production et la mise à disposition de ressources langagières utilisables bien au-delà du domaine académique. Frantext et le TLFi restent au premier plan, mais sont loin d'être les seuls produits issus des travaux de l'ATILF.

Le laboratoire continue ainsi à conforter ces études puisqu'on note de nombreuses actions de constitution et de valorisation des données (lexiques, corpus, outils permettant leur enrichissement ou leur accès). L'accès libre à ces données bénéficie à des usages en dehors du monde de la recherche et on notera parmi celles qui sont dédiées à un public identifié Lytext pour les candidats à l'épreuve orale du baccalauréat de français et Orthocorpus pour les orthophonistes.

De façon générale, il y a un effort continu des membres de l'ATILF pour la mise à disposition des produits de la recherche avec des efforts au long terme autour de Frantext qui a bénéficié d'un remaniement complet de l'outillage interne (étiquetage) et externe (interrogation). Avec cette mise à jour, les abonnements sont en augmentation.

On s'arrête sur le cas particulier d'ORTOLANG, plateforme dédiée à la diffusion des données qui permet à d'autres partenaires académiques de déposer leurs productions. Bien qu'initié dans le cadre d'un projet doté de subvention dédiées (Equipex), l'ATILF a décidé de le maintenir sur ses forces propres et le site est au final bien utilisé par les autres laboratoires français (notamment de sciences du langage).

Enfin on soulignera la proactivité du laboratoire sur le plan de l'édition ouverte.

### Points faibles et risques liés au contexte

Comme indiqué au point précédent, on regrette l'absence de partenariats avec le monde de l'entreprise, qui pourrait fournir d'autres occasions de diffuser et valoriser les ressources et outils développés dans le laboratoire.

Pour revenir sur le cas emblématique de Frantext, on soulignera sur le plan négatif le maintien des problèmes liés aux droits d'auteurs qui empêchent son ouverture totale ce qui pourrait, à terme, entraîner une désaffection. Le succès de Frantext, et les travaux importants de mise à jour pourraient sans aucun doute bénéficier d'une étude détaillée des usages, comme cela a pu se faire par le passé. Cela permettrait d'identifier les besoins et l'évolution des usagers, et ainsi anticiper les évolutions, notamment l'intégration de nouvelles ressources ou techniques permettant d'autres modes d'accès, ou de proposer des produits ou services (on peut penser aux modèles vectoriels lexicaux en tant que ressource utilisable en ingénierie linguistique, mais aussi comme façon d'enrichir les outils de navigation et d'interrogation).

Sur plusieurs sujets liés au développement et à la diffusion des ressources langagières, il est apparu un décalage entre le pilotage scientifique par les équipes thématiques et la mise en exécution par l'équipe technique. La gouvernance autour du suivi et du développement de ces ressources nécessiterait d'être précisée.

De façon générale, la prise en compte des évolutions rapides des techniques et de l'outillage en ingénierie linguistique et en TAL est vitale pour un acteur comme l'ATILF. Cela passe bien entendu par le recrutement de personnels (chercheurs et ingénieurs) capable de jouer le rôle de passeur entre les avancées récentes en informatique et en IA et les sciences du langage, pour un dialogue à double sens dans lequel l'ATILF serait à la fois producteur, consommateur et intégrateur.

### 3/ *L'unité partage ses connaissances avec le grand public et intervient dans des débats de société.*

#### Points forts et possibilités liées au contexte

On a noté dans le point précédent les actions importantes menées pour mettre à disposition de tous les produits de la recherche, et l'on sait que certains (comme le TLFi) constituent un point d'attraction pour tout type d'utilisateur.

Sur l'ensemble des activités de communication et d'animation spécifiquement destinées au grand public le rapport liste plus de 60 items.

Les membres de l'ATILF font de nombreuses interventions dans les médias, avec un large spectre de support allant de la presse quotidienne régionale à la BBC.

On notera une importante contribution au site « The conversation » dédié à la communication des chercheurs sur des sujets de société et d'actualité.

À cela s'ajoutent de nombreuses organisations ou participations à des actions ou événements, qui vont des concours d'élèves du primaire à des salons littéraires en passant par les fêtes de la science et des portraits de chercheurs.

Ces actions portent sur des sujets de sociétés variés liés au langage : orthographe, liens entre genre et langage, étymologie et histoire de la langue, enseignement des langues, etc.

Une action notable récente est la célébration (en novembre 2021) des 20 ans de l'ATILF avec une importante communication (site web, interviews, diffusion dans les médias). Le site web dédié montre d'ailleurs de façon exemplaire les compétences de l'équipe dans ce domaine, avec une grande qualité et diversité de productions.

Cette importante activité de communication repose en grande partie sur la cellule Com-valo, service interne au laboratoire qui permet un accès facilité aux médias et aux réseaux instaurés par les tutelles et les institutions locales.

#### Points faibles et risques liés au contexte

Le lien entre les sujets des interventions dans les manifestations et dans les médias avec les thèmes centraux de recherche du laboratoire n'est pas toujours évident. Il y a donc dans ce domaine un risque de dispersion des énergies à endosser un rôle de linguiste généraliste face à des sujets variés et épisodiques. Ce risque est accru par la concentration de ces actions sur un nombre réduit de chercheurs. Il serait peut-être utile de rationaliser les réponses aux demandes des médias, par exemple en s'appuyant sur la cellule Com-valo. Mais bien entendu tout repose au final sur la volonté des chercheurs eux-mêmes pour répondre à des sollicitations et à leur disponibilité.

À plusieurs reprises, le rapport (et la liste des projets financés) évoque le Projet Olki (*Open Language and Knowledge for citizens*) projet IMPACT de l'initiative Lorraine Université d'Excellence (LUE), qui concerne explicitement les liens entre les technologies du langage et les citoyens. Malgré les nombreuses mentions dans le rapport, on regrettera le manque d'information sur la structure et le rôle que l'ATILF joue en son sein. De plus, dans les nombreux projets financés dans ce cadre, on ne voit pas toujours la valeur ajoutée spécifiquement en lien avec le grand public.

## C - RECOMMANDATIONS À L'UNITÉ

### *Recommandations concernant le domaine 1 : Profil, ressources et organisation de l'unité*

Tant que les dotations pérennes ne se réorienteront pas à la hausse, l'unité sera conduite à poursuivre une implication très active dans l'obtention de projets financés, qu'ils relèvent d'instances au niveau local, national ou international. Il sera également crucial de clarifier le plus rapidement possible la situation juridique de la base Frantext afin de garantir la stabilité des recettes tirées de cette ressource.

La politique scientifique de l'unité, qui mise sur une ouverture pluridisciplinaire, ne constitue pas en soi un problème, mais il y a un intérêt patent à maintenir, voire à développer les acquis historiques afin de préserver le rôle de l'ATILF en tant que l'un des centres de gravité dans le domaine du traitement informatique de ressources dictionnaires et textuelles.

Le fonctionnement de l'unité est clairement défini et en adéquation avec les principes qui prévalent dans les structures de la recherche. Les responsabilités importantes assumées par plusieurs ITA ou BiATSS pour contribuer à l'amélioration continue du fonctionnement de l'unité, notamment en termes d'hygiène et de sécurité ou de recyclage en témoigne. Afin d'assurer la préservation du nombre de postes dévolus à l'appui à la recherche, il serait important de bien préciser la répartition et l'implication de chacun d'entre eux au sein des équipes de recherche.

### *Recommandations concernant le domaine 2 : Attractivité*

Les travaux conduits par l'unité ont fait la démonstration de leur intérêt pour le grand public (en particulier le TLFi) et pour la recherche (Démonext, PARSEME). La poursuite de programmes à longue portée (DMF, DÉRom) et le développement de nouveaux domaines (EGF, FLEURON) ne doivent pas empêcher la préservation des acquis de l'unité qui ont été mentionnés seulement de façon allusive dans le DAE : continuation du FEW, applications de la théorie Sens/Texte, maintenance d'ORTOLANG.

Une recommandation particulière concerne les ressources humaines. Alors que la démographie du corps des professeurs présage des départs importants au cours du prochain contrat, le domaine de spécialité de l'ATILF et le nombre restreint des soutenances d'HDR en interne risquent de précariser certains engagements scientifiques. Le soutien de la tutelle universitaire par un renouvellement des supports est crucial. Parallèlement, la contraction progressive des personnels IT dans l'unité, concomitante avec une réduction des crédits récurrents par le CNRS, attire l'attention sur la nécessité d'une explicitation des missions confiées aux personnels d'appui et sur les modalités de leur déploiement en interne.

S'il l'on ne peut reprocher à l'UMR l'absence de recrutement de CR ou de DR, les relations avec le CNRS gagneraient à être renforcées, au plan régional par une présence accrue au sein de la MSH Lorraine, au plan national par une collaboration plus étroite avec le consortium CORLI d'Huma-Num, avec pour perspective le montage de dossiers européens.

### *Recommandations concernant le domaine 3 : Production scientifique*

La production de l'unité est diversifiée et satisfaisante sur les plans qualitatif et quantitatif. Elle inclut une activité de collecte, de traitement, de partage et d'hébergement de données reconnue à l'international et qui bénéficie aux communautés de chercheurs et au grand public. Quelques recommandations :

- L'action auprès des tutelles doit être poursuivie dans deux directions : 1. l'obtention de recrutements du domaine des *data science* qui permettraient à ATILF de se hisser au niveau des institutions mondiales qui analysent quantitativement et modélisent les données langagières ; 2. La création d'un comité d'éthique dépendant de l'UL afin de garantir les validations éthiques nécessaires à l'accès à certaines revues et à la réponse à certains appels à projet (notamment européens).

- Les membres de l'équipe gagneraient à rendre publics leurs échanges avec la CNIL et le DPO de l'UL ainsi que les dispositions prises pour rendre leur activité conforme à la RGPD. Ils pourraient solliciter les comités opérationnels d'éthique de leurs partenaires (par exemple, le COERLE de l'INRIA) pour la validation de projets communs.

- Il revient à la direction d'engager un débat collectif sur la portée de certains constats susceptibles de freiner la propension de l'équipe à orienter sa production vers l'international (« [les] méthodologies, les discours et l'argumentation [dans d'autres langues] sont parfois très éloignés des pratiques en France et en français »).

- Enfin, la mise en œuvre d'un dispositif de suivi des doctorants, docteurs et post-doctorants, est à envisager afin que tous bénéficient d'un soutien pour valoriser et diffuser leur recherche.

## *Recommandations concernant le domaine 4 : Inscription des activités de recherche dans la société*

L'ATILF doit poursuivre ses efforts de développement et de diffusion des données auprès du public et de la communauté scientifique en favorisant le libre accès. Cela passe par le maintien de l'effort en termes de moyens accordés aux portails existants, dont Ortolang qui joue un rôle crucial dans la communauté scientifique. Il serait utile d'avoir un aperçu des usagers et des usages, afin de mieux comprendre les besoins, d'envisager les évolutions et surtout d'identifier des pistes de collaborations. Le devenir de ces ressources nécessite en interne un pilotage précis qui articule au mieux les besoins et enjeux scientifiques avec les questions plus techniques.

Concernant les partenariats avec le secteur non-académique, il est important de renforcer les activités de valorisation des ressources et des savoir-faire de l'ATILF (notamment en termes de données textuelles et lexicales) en s'appuyant sur les structures institutionnelles que l'équipe a identifiées et investies. Si les thématiques et champs sociétaux déjà traités (santé, enseignement) sont à prolonger et à approfondir, de nouvelles thématiques à envisager implique la consolidation des partenariats avec d'autres laboratoires (LORIA) ou dans le cadre des projets et outils de financement collaboratif (OLKI).

Ces efforts requièrent un renforcement des activités de la cellule « Com-valo » sur le second versant de sa mission. De même, le renforcement des compétences de l'équipe en TAL et en méthodes quantitatives (par le biais de recrutement de chercheurs et d'IT) permettrait des collaborations avec le secteur industriel.

## DÉROULEMENT DES ENTRETIENS

### DATE

**Début :** 19 septembre 2022 à 09h00

**Fin :** 19 septembre 2022 à 17h00

### Entretiens réalisés en distanciel

## PROGRAMME DES ENTRETIENS

### Le 18 sept

19h00-19h45 Réunion de travail du Comité  
Préparation de la visite en présence du Conseiller scientifique

### Le 19 sept

09h45-10h00 Vérification technique des dispositifs de visioconférence

10h00-10h30 Réunion à huis clos avec les tutelles :  
- UL : Jean-Luc Blin (Vice-président du Conseil scientifique),  
Jean-Michel Pérez (Directeur du Pôle scientifique CLCS)  
- CNRS : Ricardo Etxepare (DAS #34 Sciences du langage, INSHS),  
Edwige Helmer-Laurent (Déléguée régionale, CNRS Centre-Est)

10h30h-11h30 Session plénière  
Présentation du laboratoire : Alex Boulton, Mathieu Constant, Yan Greub, Sabrina Martin ; avec Samantha Ruvoletto

11h30-11h45 pause

11h45-12h15 Équipe Lexique  
Introduction : Alain Polguère, Veronika Lux-Pogodalla

12h15-12h45 Équipe Linguistique historique française et romane  
Introduction : Sylvie Bazin-Tacchella, Bérengère Bouard

12h45-13h15 Équipe Discours  
Introduction : Christophe Benzitoun, Samantha Ruvoletto

13h15-14h00 déjeuner

14h00-14h30 Équipe Didactique des langues et sociolinguistique  
Introduction : Maud Ciekanski, Alex Boulton

14h30-15h00 Équipe Ressources : normalisation, annotation et exploitation  
Introduction : Mathieu Constant, Véronique Montémont

15h00-15h30 Réunion à huis clos avec les doctorant.es  
Introduction : Clara Cousinard

15h30-15h45 pause

15h45-16h15 Réunion à huis clos avec les agents de soutien scientifique :  
Soutien Technique à la Recherche + Centre de Documentation M. Dinet + ITA/BIATSS des équipes scientifiques  
Introduction : Étienne Petitjean, Dominique Schloupt

16h15-16h45 Réunion à huis clos avec les agents de soutien administratif et gestion :  
Service d'Administration Générale + Communication et Valorisation  
Introduction : Sabrina Martin, Delphine Barbier

## POINTS PARTICULIERS À MENTIONNER

Pas de remarque particulière : la communication a été de bonne qualité en dépit des difficultés inhérentes aux échanges en distanciel.

## OBSERVATIONS GÉNÉRALES DES TUTELLES

**Direction de la Recherche et de  
la Valorisation**

91 avenue de la Libération  
BP454  
54001 NANCY Cedex

Clotilde BOULANGER  
[vp-recherche@univ-lorraine.fr](mailto:vp-recherche@univ-lorraine.fr)

Hélène BOULANGER  
[presidente@univ-lorraine.fr](mailto:presidente@univ-lorraine.fr)

HCERES  
2 rue Albert Einstein  
75013 Paris

Objet : Observations de portée générale sur le rapport d'évaluation - DER-PUR230022972 – ATILF (Analyse et Traitement Informatique de la Langue Française).

Madame, Monsieur,

Je vous remercie pour le rapport d'évaluation réalisé pour l'ATILF (Analyse et Traitement Informatique de la Langue Française), que vous nous avez transmis le 21 octobre 2022. Je tiens également à remercier très sincèrement les évaluateurs pour la qualité des échanges et pour l'analyse de cette unité de recherche.

Je vous prie de trouver ci-joint les observations de portée générale formulées par l'unité sur le rapport d'évaluation transmis.

Vous remerciant à nouveau pour cette évaluation qui permettra à l'unité mixte de recherche ATILF de poursuivre sa réflexion sur la base des recommandations émises, je vous prie d'agréer, Madame, Monsieur, l'expression de mes respectueuses salutations.

La Présidente de l'Université de Lorraine,  
Par délégation et pour la Présidente,  
La Vice-présidente recherche par intérim et  
Vice-Présidente en charge de la politique  
doctorale,



Clotilde BOULANGER



Nancy, le 3 novembre 2022

**Objet : Rapport Hcéres – observations de portée générale (ATILF, UMR 7118 CNRS et UL)**

Au nom de tout le laboratoire, je tiens d'abord à remercier le comité et le représentant Hcéres pour la préparation et le bon déroulement de l'évaluation et de la visite dans des circonstances difficiles après le Covid, avec la modification des consignes et des documents en cours de route. Nous avons trouvé les discussions lors de la visite très ouvertes et soulignons la valeur des commentaires dans le rapport qui montrent une profonde connaissance du DAE et des activités du laboratoire. Les nombreux points forts soulignent une activité scientifique remarquable et les « points faibles et risques liés au contexte » reflètent nos propres préoccupations ; ainsi, les recommandations s'alignent pour la plupart avec le projet quinquennal en cours de préparation qui en tiendra compte.

Nos remerciements au comité de tirer l'attention sur la **baisse de crédits** (p.6, p. 8, p.13, p.15, p.20, p.21) et la **réduction des personnels** (p.6, p.8, p.9, p.10, p.12, p.14, p.15, p.20, p.21, p.25). Le rapport insiste à de nombreuses reprises sur ces points que, de notre côté, nous soulevons régulièrement auprès des tutelles car les CDD et projets à durée limitée ne permettent pas la pérennisation des ressources qui ont fait la renommée de l'ATILF. Nous notons surtout les multiples références à la nécessité de recrutement de chercheur.es CNRS, aux départs des professeurs universitaires qui fragiliseront les travaux si des postes ne sont pas compensés, à la réduction des ITA depuis 20 ans. Le comité note en particulier les besoins en data science, statistiques, développement web.

Ce sont ces évolutions qui impactent les activités historiques du laboratoire. Les **projets lexicographiques** (p.5, p.10, p.11, p.20) dont le FEW, le DMF, le DÉRom, se poursuivent, mais à un rythme ralenti en raison du manque de personnel. Nous maintenons et développons nos **ressources d'envergure** (p.5, p.7, p.8, p.10, p.11, p.12, p.13, p.14, p.18, p.20), avec notamment des projets pour exploiter le contenu du TLFi ; Frantext est en évolution permanente et ce dispositif payant est essentiel pour les ressources propres du laboratoire ; ORTOLANG a pour objectif d'intégrer le CNRTL, son développement nécessitant un poste ETP et du matériel sur RP alors que les financements sont épuisés. Comme indiqué dans le rapport, des financements externes sont essentiels, surtout en RH ; différentes possibilités ont été explorées du côté des tutelles mais sans aboutir, « symptomatique... d'une politique nationale qui se désintéresse des programmes impulsés sans envisager leur continuité, au risque de perdre des compétences et des réalisations » (Rapport, p.13). Sans investissement des tutelles, l'ATILF ne sera bientôt plus en mesure d'assurer ces projets et devra en abandonner certains, une perte notable pour les communautés scientifiques française et internationale tout comme pour la société. Aucune réorientation ni aucun financement privé ou sur projet ne pourra combler ce déficit.

Pour les **projets internationaux** (p.11, p.13, p.20), nous avons de nombreux liens avec des organismes européens et internationaux que nous continuons à développer, avec un projet Horizon en préparation ainsi qu'un ERC *advanced grant*. Du côté du **monde économique** (p.13, p.17, p.18, p.21), l'ATILF a fait venir des représentants du SATT pour une présentation mais, comme le souligne le rapport, les possibilités sont rares dans nos domaines ; notre service Communication et Valorisation devra en effet renforcer ses activités dans ce sens. En ce qui

concerne les **partenaires locaux** (p.4, p.11, p.14, p.18, p.19, p.20), l'ATILF a maintenu des liens forts avec la MSH-Lorraine pendant une période d'instabilité de cette dernière, liens en voie de consolidation ; la MSH-L est désormais une UAR (Unité d'Appui à la Recherche). Les collaborations avec le LORIA se poursuivent pour le TAL et les ressources et seront renforcées si une double affiliation de l'équipe Sémagramme (INRIA, CNRS, UL) avec le LORIA et l'ATILF aboutit.

Au niveau des **orientations scientifiques** (p.5, p.6, p.7, p.8, p.11, p.19), sans abandonner nos activités historiques, l'ajout de nouveaux domaines (notamment psycholinguistique et didactique) est essentiel pour consolider notre place comme laboratoire en sciences du langage et n'est pas perçu comme une menace ou un dispersement des activités. En matière d'**éthique et d'intégrité scientifique** (p.16, p.20), en l'absence d'instances CNRS ou UL compétentes l'ATILF explore seul et avec la MSH-L des possibilités de créer un comité dédié. Pour les **publications en langue étrangère** (p.15, p.20), le rapport cite le DAE : « les méthodologies, les discours et l'argumentation [dans d'autres langues] sont parfois très éloignés des pratiques en France et en français. » Il convient de souligner le « parfois », car certains liens ne sont pas pénalisants (ex. dans les pays romanophones) ; l'objet était de souligner qu'il ne suffit pas toujours de traduire un texte même excellent du français vers une autre langue, l'offre institutionnelle d'aide à la traduction n'est que partiellement en adéquation. Enfin, en ce qui concerne la **protection des données** (p.9, p.20), le service STR sensibilise les membres du laboratoire aux dangers et aux obligations via une lettre d'information mensuelle ; il a mis en place un PSSI et plusieurs mesures dont le VPN, une politique de sauvegarde des données et l'achat récent de nouveaux d'onduleurs. Avec le DPD CNRS, il œuvre pour la conformité totale d'ORTOLANG au RGPD.

Il est à regretter que le manque de place accordé ne permette pas une discussion plus approfondie des nombreux autres points soulevés.

Le Directeur  
  
Alex BOULTON



Les rapports d'évaluation du Hcéres  
sont consultables en ligne : [www.hceres.fr](http://www.hceres.fr)

Évaluation des universités et des écoles

Évaluation des unités de recherche

Évaluation des formations

Évaluation des organismes nationaux de recherche

Évaluation et accréditation internationales



2 rue Albert Einstein  
75013 Paris, France  
T. 33 (0)1 55 55 60 10

[hceres.fr](http://hceres.fr)

[@Hceres\\_](https://twitter.com/Hceres_)

[Hcéres](https://www.youtube.com/Hceres)

