

## RAPPORT D'ÉVALUATION DE L'UNITÉ

LLF - Laboratoire de linguistique formelle

### SOUS TUTELLE DES ÉTABLISSEMENTS ET ORGANISMES :

Université Paris Cité

Centre national de la recherche scientifique -  
CNRS

---

**CAMPAGNE D'ÉVALUATION 2023-2024**  
VAGUE D

Rapport publié le 04/03/2024



Au nom du comité d'experts :

Alain Peyraube, président du comité

Pour le Hcéres :

Stéphane Le Bouler, président par intérim

En application des articles R. 114-15 et R. 114-10 du code de la recherche, les rapports d'évaluation établis par les comités d'experts sont signés par les présidents de ces comités et contresignés par le président du Hcéres.

Pour faciliter la lecture du document, les noms employés dans ce rapport pour désigner des fonctions, des métiers ou des responsabilités (expert, chercheur, enseignant-chercheur, professeur, maître de conférences, ingénieur, technicien, directeur, doctorant, etc.) le sont au sens générique et ont une valeur neutre.

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. Alain Peyraube, Directeur de recherche émérite, CNRS, Paris
	Mme Georgette Dal, Université de Lille
<b>Expert(e)s :</b>	Mme Julie Glikman, Université de Lorraine (représentante du CNU)
	M. Franck Sajous, CNRS, Toulouse (personnel d'appui à la recherche)
	Mme Efstathia Soroli, Université de Lille (représentante du CoNRS)

## REPRÉSENTANT DU HCÉRES

M. Bernard Laks

## REPRÉSENTANTS DES ÉTABLISSEMENTS ET ORGANISMES TUTELLES DE L'UNITÉ DE RECHERCHE

M. Ricardo Extépare, INSHS, CNRS  
Mme Marie-Hélène Papillon, Délégation régionale CNRS  
Mme Marie Salaün, faculté Sociétés et Humanités, Université Paris Cité

## CARACTÉRISATION DE L'UNITÉ

- Nom : Laboratoire de Linguistique Formelle
- Acronyme : LLF
- Label et numéro : UMR 7110
- Composition de l'équipe de direction : M. Olivier Bonami (directeur) et Mme Giuseppina Turco (directrice adjointe)

## PANELS SCIENTIFIQUES DE L'UNITÉ

SHS Sciences humaines et sociales

SHS4 L'esprit humain et sa complexité

ST6 Sciences et technologies de l'information et de la communication

Le Laboratoire de linguistique formelle LLF est une unité mixte de recherche du CNRS et de l'université Paris Cité, qui relève à la fois du domaine des sciences humaines et sociales (SHS) et du panel SH4, L'esprit humain et sa complexité, et du domaine des sciences et technologies de l'information et de la communication (STIC) et du panel ST6.

## THÉMATIQUES DE L'UNITÉ

L'objectif de long terme du LLF est l'étude intégrée du langage à travers les langues. Le langage est étudié sous tous ses aspects : phonétique, phonologie, morphologie, syntaxe, sémantique, pragmatique, sans oublier les dimensions diachronique, sociolinguistique ou psycholinguistique ni les applications en didactique et en traitement automatique des langues.

Aussi, depuis sa fondation le LLF a-t-il choisi de ne pas avoir de structuration en équipe. Il a adopté un fonctionnement souple où s'entrecroisent quatre thèmes de recherche et quatre axes transversaux, qui favorisent les collaborations entre spécialistes de différentes sous-disciplines. Chaque thème ou axe transversal est placé sous la responsabilité d'un chercheur qui l'anime et informe la direction et le conseil des activités qui y sont menées. Cette structuration à deux entrées permet d'encourager les collaborations autour de domaines ou de méthodes partagées, avec une grande souplesse, en fonction d'une évolution rapide de l'effectif, de ses compétences et de ses intérêts.

Les quatre domaines thématiques sont les suivants : le mot, la phrase, discours et logique, variation et changement. Les quatre axes regroupent la linguistique expérimentale, la linguistique computationnelle, la linguistique descriptive et de terrain, les formalismes linguistiques.

Le domaine du « mot » regroupe l'ensemble des recherches menées au sein du laboratoire portant sur la structure et l'interprétation des objets lexicaux : morphosyntaxe, sémantique lexicale, morphologie flexionnelle et dérivationnelle, typologie des systèmes morphologiques. Les principales questions de recherche abordées sont les travaux portant sur les limites de l'identité lexicale, ainsi que les études sur la typologie et la modélisation des systèmes morphologiques.

La thématique sur « la phrase » est surtout consacrée aux aspects syntaxiques et sémantiques. Elle comprend des opérations sur les relatives et leur traitement psycholinguistique, sur la pluralité et la quantification, sur la négation, sur les relations entre actif et passif.

Le domaine « discours et dialogue » mobilise des études de corpus et des techniques expérimentales appliquées à plusieurs langues pour répondre à des questions de recherche portant sur le discours et le dialogue, qui mobilisent des implémentations computationnelles.

Enfin le domaine thématique « variation et changement », mis en place plus récemment, porte sur l'étude des langues naturelles dans leurs aspects dynamiques, à savoir comment leurs propriétés varient selon les propriétés sociales, historiques et développementales de leurs locuteurs. Les recherches dans ce domaine sont caractérisées d'un côté par leur large empan empirique (langues romanes, langues créoles, langues de l'Asie de l'Est, entre autres) et, de l'autre, par leurs méthodes innovantes, notamment à travers les expériences psycholinguistiques et la modélisation grammaticale formelle et computationnelle. Ce domaine s'articule en quatre opérations : variation diachronique, langues créoles, l'acquisition des L1 et L2, la sociolinguistique.

## HISTORIQUE ET LOCALISATION GÉOGRAPHIQUE DE L'UNITÉ

Le LLF a été fondé en 1972 pour être un centre de recherches en linguistique à l'université Paris 7. Dès les années 1990, le LLF s'est développé autour d'une approche résolument formelle visant une modélisation mathématique des phénomènes et systèmes linguistiques. Le LLF est alors devenu en 2001 une UMR avec pour tutelles l'université Paris 7 et le CNRS. Au début des années 2010, l'unité s'est ouverte à de nouvelles méthodologies, en particulier la linguistique expérimentale et les études quantitatives de corpus, en mobilisant un ensemble plus large de sources de données, allant des résultats d'expériences psycholinguistiques à des données attestées, quantifiées. À partir de 2017, l'arrivée de quatre nouveaux membres travaillant en linguistique computationnelle a renforcé cette thématique au sein du laboratoire. Ceci a entraîné une structuration en grille pour la période 2017-2023 comportant les quatre axes transversaux et les quatre domaines détaillés ci-dessus.

L'unité, initialement localisée sur le campus de Jussieu, dans le 5<sup>e</sup> arrondissement de Paris, a connu une décennie de déménagements, notamment rue du Château-des-Rentiers, dans le 13<sup>e</sup> arrondissement, avant de rejoindre en 2013 ses locaux actuels, dans le bâtiment Olympe de Gougues du campus Paris Rive Gauche de l'université Paris Cité.

Le LLF y dispose d'un espace de 580 m<sup>2</sup> de bureaux aux 5<sup>e</sup> et 6<sup>e</sup> étages, au sein de l'UFR de linguistique de l'université, où il est également au contact direct des autres unités de recherche de l'université (HTL – Laboratoire d'histoire des théories linguistiques, UMR 7597, CLILLAC-ARP – Centre de linguistique interlangue, de lexicologie, de linguistique anglaise et de corpus, Atelier de recherche sur la parole, UR 3967) ainsi que de la composante de l'université Paris Cité de l'École Doctorale (ED 622) de Sciences du langage. Dans ce bâtiment Olympe de Gougues, le LLF profite également de locaux mutualisés, et notamment d'une salle d'expérimentation et d'une salle serveurs.

## ENVIRONNEMENT DE RECHERCHE DE L'UNITÉ

Le LLF, qui inscrit ses activités de recherche au sein de l'UFR de linguistique de la Faculté « Sociétés et Humanités » de l'université Paris Cité, travaille en étroite collaboration avec l'École Doctorale 622 Sciences du langage, qui relève à la fois de l'université Paris Cité et de l'université Sorbonne Nouvelle, pour encadrer ses doctorants et doctorantes.

Il est aussi membre fondateur du Labex (Laboratoire d'excellence) EFL (Empirical Foundations of Linguistics, 2011-2024), et en assure la direction. Il héberge dans ses locaux le personnel de soutien qui lui est rattaché. L'unité est aussi un contributeur décisif à l'EUR (École universitaire de recherche) PGS (Paris Graduate School of Linguistics) qui rassemble, sur le modèle international reconnu, des formations de master et de doctorat d'excellence.

L'unité bénéficie en outre, par des appels à projets, de divers financements de l'idex (Initiative d'excellence) de l'université Paris Cité, notamment une chaire d'excellence « Experimental investigation of discourse connectives », des projets Émergence en recherche, etc.

Le laboratoire participe également à l'activité et au pilotage de l'unité interdisciplinaire DIIP (Data Intelligence Institute of Paris), qui rassemble des chercheurs dans les sciences de la matière et de la vie, ainsi que dans le domaine des sciences humaines et sociales.

Le LLF bénéficie enfin des dispositifs de réseaux internationaux du CNRS : GDRI SeePiCLa (Groupe de Recherche International Structure, Émergence and Evolution of Pidgin and Creole Languages, 2012-2020) ; LIA Ro-GraV (Laboratoire International Associé « The Grammar of the Verbal Domain in Romanian and Beyond », 2017-2021) ; IRP (International Research Project FEMIDAL « (Formal/Experimental Methods and In-depth Description of Australian Indigenous Languages », 2021-2026).

## EFFECTIFS DE L'UNITÉ : en personnes physiques au 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	7
Maitres de conférences et assimilés	12
Directeurs de recherche et assimilés	3
Chargés de recherche et assimilés	5
Personnels d'appui à la recherche	6
<b>Sous-total personnels permanents en activité</b>	<b>33</b>

Enseignants-chercheurs et chercheurs non permanents et assimilés	9
Personnels d'appui non permanents	3
Post-doctorants	3
Doctorants	29
<b>Sous-total personnels non permanents en activité</b>	<b>44</b>
<b>Total personnels</b>	<b>77</b>

RÉPARTITION DES PERMANENTS DE L'UNITÉ PAR EMPLOYEUR : en personnes physiques au 31/12/2022. Les employeurs non tutelles sont regroupés sous l'intitulé « autres ».

Nom de l'employeur	EC	C	PAR
Université Paris Cité	16	0	2
CNRS	0	8	4
Autres	3	0	0
<b>Total personnels</b>	<b>19</b>	<b>8</b>	<b>6</b>

## AVIS GLOBAL

Le LLF a pour objectif essentiel l'étude intégrée du langage à travers les langues. Le langage y est étudié sous tous ses aspects : phonétique, phonologie, morphologie, syntaxe, sémantique, pragmatique, sans oublier les dimensions diachronique, sociolinguistique ou psycholinguistique ni les applications en didactique et en traitement automatique des langues. C'est dire que presque aucun domaine ou sous-domaine des sciences du langage n'est laissé de côté.

Cette stratégie a permis à LLF d'occuper une place nationale et internationale reconnue de premier rang en sciences du langage, notamment pour avoir su prendre avec succès le tournant empirique de la théorie que l'on peut observer depuis une vingtaine d'années, mieux sans doute que d'autres unités de recherche en linguistique dans le dispositif de recherche français. Pour ce faire, les recherches menées dans l'unité se sont enrichies de méthodes quantitatives, expérimentales et computationnelles. Cette articulation est bien reflétée dans la structuration à double entrée adoptée par le LLF, qui croise les quatre domaines thématiques scientifiques de l'unité et les quatre axes transversaux : linguistique expérimentale, linguistique computationnelle, linguistique descriptive et de terrain, ainsi que les formalismes linguistiques.

L'attractivité du laboratoire est aussi attestée par le volume important de ressources financières propres qui complètent efficacement les dotations de base des tutelles : 85 % du budget du laboratoire est issu de ressources propres. Le LLF sait tirer profit de son environnement de recherche et des compétences indiscutables de ses chercheurs et EC (enseignants-chercheurs) pour répondre avec succès à toutes sortes d' AAP (appels à projets) compétitifs, aux niveaux local, national et, dans un moindre degré, européen.

Outre les projets inclus dans le labex EFL dont l'unité assure la direction, l'unité bénéficie d'un ERC (European Research Council) Starting grant (Formal Models of Social Meaning and Identity Construction through Language 2020-2025) et d'un projet H-2020 (Horizon-2020) intitulé 'The Sign Hub: Preserving, Researching and Fostering the Linguistic, Historical and Cultural Heritage of European Deaf Signing Communities with an Integral Resource, 2016-2020'. S'y ajoutent une chaire d'excellence (Experimental investigations of discourse connectives), cinq projets nationaux ANR en cours (Analyser les modèles d'apprentissage profond pour extraire les informations typologiques d'une langue, 2024-2028 ; Biais inductifs pour des modèles compositionnels du langage naturel 2024-2028 ; Mathematical Hygiene, 2024-2028 ; Sentential Negation Across the Globe: The fine-grained structure of a linguistic universal, 2023-2027) ; et la participation à un projet financé par la National Research Foundation of Korea SELEXIN (SEmantic LEXicon INduction for Interpretability and diversity in text processing, 2022-2026). Cette capacité de l'unité à obtenir des financements en réponse à des appels à projets compétitifs est sur une trajectoire ascendante, même si des marges de progression demeurent au niveau européen.

Ces succès ont aussi contribué financièrement à l'équipement de deux plateformes appuyant des recherches dans les domaines clés de la linguistique expérimentale et de la linguistique informatique. Cet équipement est composé actuellement de deux oculomètres fixes, d'un système de lunettes oculométriques, d'un système d'électro-encéphalographie, d'un système de captation de la conductivité de la peau et un échographe. Le LLF administre et utilise également une plateforme de linguistique informatique, qui comporte plusieurs serveurs de calcul et de stockage, un serveur dédié aux expériences psycholinguistiques en ligne et un dédié à la démonstration des ressources développées au laboratoire. Le tout représente un investissement de plus de 80 k€.

La production scientifique de l'unité est cohérente avec sa taille, ses activités et tous ses domaines d'intervention. Elle est remarquable par sa quantité et sa qualité, sur des supports de publication internationaux et reconnus. Il est vrai que la politique du laboratoire incite fortement ses personnels à publier, y compris les PAR (personnels d'appui à la recherche), les étudiants dès le premier niveau (Master) et les post-doctorants. Les supports sont de qualité et de haute visibilité internationale : Language, Journal of linguistics, Morphology. Dans les domaines de la linguistique computationnelle et de la phonétique/phonologie on note la participation et la publication dans les actes des grandes conférences internationales, véritables références mondiales sélectives des domaines : LREC (Language Resources and Evaluation Conference), COLING (International Conference on Computational Linguistics), ACL (Association of Computational Linguistics), EMNLP (Conference on Empirical Methods in Natural Language Processing, Interspeech, SpeechProsody). Le comité a relevé cependant que certains chercheurs restent moins identifiés au-delà de la communauté scientifique nationale que d'autres. Il recommande en conséquence de mettre en place des dispositifs favorisant la visibilité et le rayonnement scientifique de tous ses membres, par exemple en proposant des aides à la traduction, une liste de revues ciblées comme prioritaires dans différents domaines, en promouvant la co-publication.

L'unité est par ailleurs particulièrement attentive à la formation à la recherche par la recherche. Elle participe pleinement aux enseignements en licence, en master, à l'ED 662 « Sciences du langage » de Paris Cité et dans les deux Graduate Schools récemment créées, la PGSL et la GRAIDS (Graduate School Artificial Intelligence and Data Science).

En ce qui concerne la communication, l'unité a une stratégie assez hétérogène. Les séminaires, formations, et nouvelles productions de l'unité circulent principalement sur les listes de diffusion du laboratoire mais elles sont rarement publiques. Le comité encourage l'unité à investir davantage dans la communication et la diffusion de ses recherches, événements organisés, formations, etc.

Pour ce qui est enfin de la publication des sources et des données de la recherche, le DAE, pas plus que la visite de l'unité, n'ont permis de mesurer la politique mise en place et ses effets : aucune indication précise sur les publications de codes sources, ressources et corpus n'a été fournie. Le comité attire l'attention du laboratoire sur le fait que dans une politique de science ouverte, la publication des données de la recherche (codes sources, ressources, corpus, données) est tout aussi centrale que la publication de contributions scientifiques résultant de l'analyse de ces données.

# ÉVALUATION DÉTAILLÉE DE L'UNITÉ

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

Le précédent rapport avait recommandé de :

- 1, Développer les travaux en psycholinguistique et en neurolinguistique ;
- 2, Augmenter la visibilité des productions, en favorisant les publications dans des revues ;
- 3, Améliorer la visibilité des ressources électroniques produites par l'unité en s'appuyant sur les infrastructures nationales, notamment Huma-Num ;
- 4, Obtenir de nouveaux financements externes ;
- 5, Améliorer l'impact sociétal des recherches menées dans l'unité ;
- 6, Réduire la durée des préparations de thèses ;
- 7, Équilibrer la taille des domaines thématiques, notamment en termes d'encadrement doctoral ;
- 8, Formaliser le règlement intérieur du laboratoire.

La quasi-totalité de ces recommandations ont été prises en considération par la direction de l'unité :

1/ L'unité a continué à développer son implication dans les recherches en linguistique expérimentale à partir de données comportementales ou physiologiques. 18 des 46 thèses soutenues dans la période ont donné lieu à des travaux expérimentaux. L'activité dans le domaine « discours et dialogue » a été particulièrement importante, notamment grâce au recrutement d'un directeur de recherche du CNRS, relevant de la section 26.

2/ Le conseil de laboratoire n'a pas décidé de mettre en place une authentique politique d'incitation active pour les publications, mais a beaucoup communiqué auprès des membres de l'unité sur l'importance du choix des supports. La part des publications d'articles dans des revues est passée de 27 % à 38 %, alors que celle des communications sans actes dans le total des productions est passée de 39 % à 23 %.

3/ Le LLF s'appuie sur diverses infrastructures nationales dont Huma-Num lorsque cela est pertinent. L'unité participe activement au pilotage et aux actions du consortium Corli (Corpus Langues et Interactions). Cependant, la plupart des ressources produites au LLF ayant une vocation internationale et non nécessairement francophone sont disponibles sur des plateformes et des collections ayant une meilleure visibilité pour les utilisateurs (sites locaux dédiés, Open Science Foundation, Zenodo, CLARIN [Common Language Resources and Technology Infrastructure], etc.).

4/ La précédente évaluation a eu lieu dans une période d'incertitude à propos du renouvellement du Labex EFL. Cette incertitude a été levée. En outre, l'unité a continué une forte politique d'incitation au dépôt de projets, avec un succès très net : la part des ressources propres est passée de 73 % pour le précédent contrat à 85 % pour celui-ci, et les sources de financement se sont encore diversifiées (Idex, ANR, ERC, région, Ministère des Armées, contrats industriels).

5/ La direction de l'unité continue d'encourager toutes les actions de valorisation et de dissémination qui font sens au vu de ses thématiques de recherche. Sur la période, on peut noter l'augmentation du nombre de thèses dans le cadre du dispositif Cifre (Conventions industrielles de formation par la recherche) ou assimilées, qui est passé d'un à quatre dans le domaine du traitement automatique des langues, ainsi que diverses actions prospectives dans les domaines de la didactique des langues et des politiques linguistiques qui n'ont pas encore abouti à ce jour. L'unité mène en outre des actions très innovantes pour la santé des populations aborigènes d'Australie. Dans le domaine de la dissémination, l'unité a eu une activité très remarquable autour de la sortie de la GGF, ainsi que dans les débats sur l'utilisation du genre grammatical en français.

6/ Le laboratoire poursuit la politique entamée lors du précédent contrat, dont les effets ont mis du temps à se faire ressentir, et qui consiste à limiter le nombre de thèses en se concentrant sur des encadrements dans les meilleures conditions, avec un financement complet. Les effets positifs de cette politique sont peu visibles du fait des ralentissements dus au Covid-19, qui ont justifié une prolongation des contrats doctoraux de 6 mois en moyenne pour les thèses inscrites en 2018 et 2019. Malgré cela, on note que la durée moyenne des thèses est passée de quatre ans onze mois à quatre ans huit mois, et la médiane de quatre ans quatre mois à quatre ans deux mois depuis le précédent contrat.

7/ Les recommandations concernant l'organisation ont été prises en compte dans la réflexion sur la structuration de l'unité. Les quatre nouveaux domaines méthodologiques sont mieux équilibrés.

8/ En matière de règlement intérieur, les efforts entrepris à ce sujet dès 2018 se sont heurtés à l'instabilité des personnels de support juridique des tutelles du fait de la création de l'université Paris Cité et de la crise de la Covid-19. Le règlement intérieur est en passe d'être adopté par le conseil.

## B - DOMAINES D'ÉVALUATION

Consigne de rédaction pour tous les domaines d'évaluation (1, 2, 3 et 4) : En considérant les références définies dans le référentiel d'évaluation des UR, le comité veille à distinguer les éléments remarquables, qui se rapportent à des points forts ou à des points faibles. Chacun des points est étayé par des faits observables notamment à partir des éléments déposés dans le portfolio. Le comité apprécie si le bilan de l'unité est en cohérence avec son profil d'activités.

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

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

Le LLF occupe une place nationale et internationale reconnue de premier rang dans le domaine des sciences du langage et a pris avec succès le tournant empirique de la théorie. L'unité s'inscrit pleinement dans l'écosystème local, national et international, et prend une part très active dans les structures de recherche existantes. Elle développe une politique d'incitation à la publication auprès de ses ingénieurs. Elle est par ailleurs très engagée dans la formation à la recherche par la recherche, que ce soit en licence, en Master, à l'ED 662 « Sciences du langage » ou dans les deux « Graduate Schools » récemment créées.

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

Les ressources financières sont particulièrement favorables, et le LLF est très actif dans le développement de ses ressources propres, aussi bien au niveau des projets publics de site, nationaux et internationaux qu'au niveau des projets relevant du privé. L'unité met à disposition de ses chercheurs, enseignants-chercheurs et ingénieurs deux plateaux techniques, en accord avec l'objectif scientifique qu'elle s'est assigné. En matière de ressources humaines et patrimoniales, ses succès répétés dans les réponses à appels ont cependant comme corollaire un manque de personnels d'appui à la recherche ainsi qu'un manque d'espaces susceptibles d'accueillir dans de bonnes conditions des personnels non permanents.

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

Le fonctionnement de l'unité apparaît satisfaisant, et respecte les réglementations en matière de gestion des ressources humaines, de sécurité, d'environnement et de protection du patrimoine scientifique, dans la limite de ses champs de compétences, dans la mesure où elle est en grande partie tributaire de la politique de ses tutelles en la matière. La part croissante de personnels non francophones a pour effet que la langue anglaise est devenue la langue de communication scientifique de l'unité par défaut, ce qui est source de difficulté pour les personnes qui ne sont pas à l'aise dans cette langue. Le règlement intérieur du LLF doit encore être adopté formellement. L'unité ne s'est, pour l'heure, fixé aucun objectif de développement durable.

*1/ L'unité s'est assigné des objectifs scientifiques pertinents.*

#### Points forts et possibilités liées au contexte

Le LLF s'est assigné comme objectif d'être un acteur de rang mondial en linguistique théorique et descriptive, en phase avec le tournant empirique de la théorie que l'on peut observer depuis une vingtaine d'années. Pour ce faire, les recherches menées dans l'unité s'enrichissent de méthodes quantitatives, expérimentales et computationnelles. Cette articulation est reflétée dans la structuration à double entrée adoptée par le LLF, qui croise les quatre domaines thématiques scientifiques de l'unité et les quatre axes transversaux, que constituent la linguistique expérimentale, la linguistique computationnelle, la linguistique descriptive et de terrain ainsi que les formalismes linguistiques.

Le dynamisme et la visibilité des travaux menés dans chacun de ces domaines et de ces axes témoignent d'un niveau comparable d'excellence scientifique.

Les projets de l'unité s'inscrivent en outre pleinement dans l'écosystème local, national et international, comme l'attestent le Labex EFL dont le LLF est membre fondateur et qu'il copilote, l'EUR PGSL que dirige l'un des membres de l'unité, l'obtention de divers projets au sein de l'Idex de l'université Paris Cité, dont la chaire d'excellence « Experimental investigation of discourse connectives », la participation à l'activité et au pilotage de l'institut interdisciplinaire « Data Intelligence Institut of Paris », ainsi que la participation à huit projets ANR (Agence Nationale de la Recherche), dont trois portés. Il s'agit des projets DUEL (« Dysfluences Utterances, Exclamations and Laughter in Dialogue »), GEOMPHON (« Speech Perception and Learning in the Geometric Typology of Phonological Inventories ») et SNAG (« La négation phrasique à travers le monde »). Pendant la période, le LLF a également participé à plusieurs réseaux internationaux du CNRS : le GDRI SeePiCLa (Groupement de Recherche International « Structure, Emergence and Evolution of Pidgin and Creole Languages », 2012-2020), le LIA Ro-GraV (Laboratoire International Associé « The Grammar of the Verbal Domain in Romanian and Beyond », 2017-202) ainsi que l'IRP FEMIDAL (International Research Project « Formal/Experimental Methods and in-depth Description of Australian indigenous Languages », 2021-2026).

L'unité est par ailleurs particulièrement attentive à la formation à la recherche par la recherche. Elle participe pleinement aux enseignements en licence, en master, à l'ED 662 « Sciences du langage » et dans les deux Graduate Schools récemment créées, la PGSL et la GRAIDS. S'agissant du niveau master, outre l'accueil d'étudiants pour des stages longs, on note sept publications issues d'un mémoire de master, dont l'étudiant est premier auteur.

Les ingénieurs du LLF sont également fortement impliqués dans l'activité de publication et de production de ressources, et ont cosigné dix-sept publications sur la période.

### Points faibles et risques liés au contexte

Pas de point faible identifié.

*2/ L'unité dispose de ressources adaptées à son profil d'activités et à son environnement de recherche et les mobilise.*

### Points forts et possibilités liées au contexte

Le point fort remarquable du LLF est sa capacité à tirer profit de son environnement de recherche et de développer des ressources propres en répondant à des appels à projets compétitifs. Pendant la période, le LLF a ainsi répondu avec succès à (i) un contrat international avec le MIT (Massachusetts Institut of Technology) : « Evaluating meaning based approach to island constraints », dont il est porteur ; (ii) deux contrats européens, dont le projet SMIC « Formal Models of Social Meaning and Identity Construction through Language », porté par l'unité et financé par l'ERC ; (iii) huit contrats ANR, dont trois portés (cf. *supra*) ; (iv) un contrat porté CNRS-ministère des armées ; (v) cinq projets dans le cadre du PIA, tous portés, dont la chaire d'excellence « Experimental investigations of discourse connectives ». En outre, cinq financements ont été obtenus dans le cadre de partenariats avec le monde socio-économique et culturel, tous portés, dont trois thèses relevant du dispositif Cifre, un contrat avec une collectivité territoriale et un contrat avec une fondation. Selon les années, ces ressources propres ont représenté entre trois quarts et huit neuvièmes du budget de l'unité, les dotations récurrentes reçues des tutelles étant, quant à elles, restées stables.

Le second point fort de l'unité réside dans les deux plateaux techniques dont elle dispose : la plateforme de linguistique expérimentale, partagée avec une autre unité de recherche, ainsi que la plateforme de linguistique informatique, qui lui est propre.

### Points faibles et risques liés au contexte

Sur la période, le LLF a bénéficié de la part du CNRS d'un IE (Ingénieur d'Études) et d'un AI (Assistant Ingénieur) supplémentaires, portant ainsi le nombre de personnels d'appui à la recherche, gestionnaires inclus, à 5,5.

Ces personnels peinent cependant à faire face aux besoins croissants en ingénierie scientifique et en calcul, requis par les plateformes qu'administre le LLF, ainsi qu'en matière de gestion RH et financière pour ce qui relève de la gestion des ressources propres et de l'accueil de stagiaires rémunérés. Sur le plan patrimonial, l'espace de bureaux attribués au LLF dans l'UFR de linguistique de l'université Paris Cité est saturé : il dispose de 67 postes de travail, pour une moyenne annuelle de 70 membres, personnels non permanents inclus, auxquels s'ajoutent l'accueil d'une dizaine de visiteurs par an, ainsi que celui d'une trentaine de stagiaires en Master. La politique

de formation à la recherche par la recherche ainsi que la dynamique de réponses à appel à projets assortie du recrutement de contractuels peuvent ainsi se trouver entravées.

### *3/ Les pratiques de l'unité sont conformes aux règles et aux directives définies par ses tutelles en matière de gestion des ressources humaines, de sécurité, d'environnement, de protocoles éthiques et de protection des données ainsi que du patrimoine scientifique.*

#### Points forts et possibilités liées au contexte

Les 33 membres permanents de l'unité au 31/12/2022 se répartissent en 54,5 % de femmes et 45,5 % d'hommes. La parité est par conséquent respectée. À un grain plus fin, elle l'est strictement chez les chercheurs, avec quatre femmes et quatre hommes, un peu moins chez les enseignants-chercheurs, avec douze femmes (63 %) et sept hommes (37 %). Ce déséquilibre est dû à une sur-représentation des femmes dans le corps des maîtres de conférences, qui compte huit femmes (67 %) pour quatre hommes (33 %). Même si la marge de manœuvre est limitée compte tenu du faible nombre d'entrées-sorties d'enseignants-chercheurs sur la période, l'unité porte par conséquent une attention à la question de la parité genrée, qu'elle applique également dans l'attribution de responsabilités et la composition du conseil d'unité.

Les conditions de travail des personnels permanents sont bonnes, chacun bénéficiant d'un poste de travail attitré. Les permanents et les doctorants disposent en outre d'un ordinateur portable professionnel.

L'un des membres permanents du LLF est assistant de prévention.

Un groupe de travail a été lancé au sein de l'unité sur la question des risques psychosociaux.

La protection des systèmes informatiques est du ressort de la tutelle de l'université Paris Cité. Le LLF a cependant mis en place plusieurs mesures de protections individuelles : par exemple, sensibilisation au phishing et au ransomware, incitation à la sauvegarde des données de la recherche sur des plateformes institutionnelles, chiffrement systématique des disques durs.

#### Points faibles et risques liés au contexte

La part croissante de personnels non francophones a pour effet que la langue anglaise est devenue la langue de communication scientifique par défaut, ce qui est source de difficulté pour les personnes qui ne sont pas à l'aise dans cette langue.

Le règlement intérieur du LLF, dont l'absence a été soulignée lors de la dernière évaluation par le Hcéres, doit encore être formellement adopté. Si un travail a été engagé sur ce point dès 2018, la création de l'université Paris Cité à laquelle s'est greffé le ralentissement des activités pendant la crise sanitaire n'a pas permis à ce règlement d'être voté.

L'unité ne s'est, à ce jour, fixé aucun objectif de développement durable, que ce soit en matière de déplacements ou d'achats responsables.

La prévention des risques psychosociaux ne semble pas avoir été l'objet d'une réflexion aboutie, particulièrement en ce qui concerne les questions de harcèlement.

## DOMAINE 2 : ATTRACTIVITÉ

### Appréciation sur l'attractivité de l'unité

Le rayonnement scientifique de l'unité est important et ses manifestations sont nombreuses. Le LLF se positionne comme un acteur de poids au niveau international en linguistique théorique et empirique. Ses membres s'impliquent fortement dans l'animation scientifique avec l'organisation et la participation à de nombreux événements scientifiques, avec la prise de responsabilités éditoriales et avec un investissement accru dans le pilotage de la recherche en France. Le LLF mène une activité soutenue de réponse à des appels à projets locaux, régionaux, nationaux et européens, ce qui a conduit à un doublement des ressources propres qui constituent maintenant 85 % de son budget. Ces succès ont contribué à lui permettre de se munir de deux plateformes appuyant des recherches dans les domaines clés de la linguistique expérimentale et de la linguistique informatique. La capacité de l'unité à obtenir des financements par des appels à projets compétitifs est de fait sur une trajectoire ascendante, même si des marges demeurent au niveau européen.

- 1/ L'unité est attractive par son rayonnement scientifique et s'insère dans l'espace européen de la recherche.
- 2/ L'unité est attractive par la qualité de sa politique d'accompagnement des personnels.
- 3/ L'unité est attractive par la reconnaissance de ses succès à des appels à projets compétitifs.
- 4/ L'unité est attractive par la qualité de ses équipements et de ses compétences techniques.

#### Points forts et possibilités liées au contexte pour les quatre références ci-dessus

Le rayonnement scientifique de l'unité est évident et ses manifestations sont nombreuses pendant la période de référence. Malgré la pandémie, l'unité a été impliquée dans l'organisation de 31 colloques ou ateliers de recherche : seize sur place à Paris, sept à l'étranger, huit en ligne, dont plusieurs de grande envergure. Le LLF a également organisé ou coorganisé quatre écoles d'été. Les membres du LLF sont aussi très actifs dans des activités éditoriales telles que la direction de revues (comme *Morphology, Dialogue and Discourse*), la participation à des comités de rédaction (par ex. *Cahiers de lexicologie, Études Créoles, Faits de langue, Isogloss, Journal of Semantics, Semantics and Pragmatics*), la direction de séries dans des grandes maisons d'édition internationales comme Springer et Language Science Press, ou encore la participation à des comités éditoriaux de séries chez Brill, Edinburgh University Press, Language Science Press, Routledge.

Pendant la période d'évaluation, nombreux ont été les membres du LLF qui ont siégé dans des instances de pilotage de la recherche : dans la section 34 du Comité national de la recherche scientifique (trois), dans le cadre du CNU (trois). Des membres du laboratoire ont aussi participé à des comités scientifiques de l'ANR, à des jurys de l'Institut Universitaire de France, à des évaluations d'unités ou de projets de recherche pour le Hcéres et l'agence allemande DFG (Deutsche Forschungsgemeinschaft). Enfin, le LLF a participé à trois dispositifs de réseaux internationaux du CNRS : GDRI SeePiCLa (2012-2020), LIA Ro-GraV (2017-2021), IRP FEMIDAL (2021-2026), pilotés par l'unité.

Dans le cadre des concours du CNRS pour les chargés de recherche, le LLF lance des appels à candidatures ouverts, aide les candidats à préparer leur dossier (y compris en les auditionnant) et présélectionne les meilleures candidatures. Cette politique permet d'attirer des jeunes chercheurs internationaux et a conduit à des recrutements. Après le recrutement, la direction porte une attention particulière aux conditions d'accueil pour offrir des bonnes conditions matérielles et les informations nécessaires sur le fonctionnement de l'unité, proposant des échanges réguliers pour coordonner de manière informelle un rythme de prises d'initiative et de responsabilité compatible avec la situation des personnes recrutées. La direction joue également un rôle de soutien pour la promotion des chargés de recherche et des maîtres de conférences en mettant à leur disposition un plafond de dépenses pour des missions plus élevé que celui qui est dévolu aux directeurs de recherche et professeurs, afin de les encourager à enrichir rapidement leur liste de publication et leur réseau. Le LLF héberge enfin un nombre important d'émérites (neuf, soit l'équivalent d'un tiers de l'effectif permanent), pour la plupart encore très actifs, qui bénéficient, dans la mesure du possible, du même traitement que les membres permanents.

Le LLF s'appuie sur les financements du labex EFL de l'université Paris Cité, et sur ses ressources propres pour mener une politique ambitieuse et attirer des chercheurs étrangers. Sur la période d'évaluation, le laboratoire a accueilli 21 visiteurs étrangers pour des séjours d'au moins deux semaines, malgré l'épidémie de Covid-19. En parallèle, dans la période de référence, la direction et les membres de l'unité ont été soucieux d'améliorer les conditions d'accueil et de travail des doctorants. Plus précisément, le laboratoire a envoyé seize doctorants pour des séjours de recherche de trois mois dans des laboratoires étrangers et accueilli un nombre équivalent de jeunes chercheurs étrangers.

Sur la période, le laboratoire a enregistré 37 inscriptions en thèse et a accueilli dix-sept postdoctorants. Le LLF a pour politique de soutenir toutes les demandes d'inscription en thèse faisant l'objet d'un financement, mais examine au cas par cas les éventuelles demandes sans financement. Le comité note qu'aucune inscription sans financement n'a eu lieu pendant ce contrat.

La renommée du laboratoire se mesure enfin à des reconnaissances explicites extérieures comme un doctorat Honoris Causa de l'université Charles de Prague (CZ), une nomination à l'Academia Europaea, et quatre nominations à l'Institut Universitaire de France.

La direction du laboratoire mène une veille active sur les appels à projets, et informe les membres de l'unité des opportunités. Dans les cas où le laboratoire ne peut soutenir qu'un nombre limité de projets ou doit les classer, une procédure d'arbitrage transparente est mise en place par le Conseil. Sur la période, on note 27 projets obtenus à la suite d'appels à projets compétitifs, ce qui a permis d'une part la structuration et la visibilité internationale d'une communauté de recherche d'excellence au niveau régional, et d'autre part d'asseoir le positionnement interdisciplinaire du LLF au sein de la nouvelle université.

Les projets obtenus se répartissent comme suit : deux projets européens, un ERC Starting grant et un H-2020 ; neuf projets nationaux dont huit projets ANR ; un projet relevant du PIA, le labex EFL ; dix projets internes à l'université Paris Cité, certains financés par l'Idex ; deux projets de réseaux internationaux du CNRS ; un projet financé par la Région Île-de-France ; un projet financé par une entreprise privée ; un projet financé par la fondation Fyssen. Ces projets permettent également le recrutement de personnels non permanents. Ils ont financé : dix-sept contrats de postdoctorats pour un total de 266 personne-mois ; quinze contrats doctoraux, soit 20 % des thèses soutenues ou préparées dans l'unité ; six recrutements de PAR ; ainsi que l'équipement des deux plateformes du laboratoire, à hauteur de 50 k€ environ.

Le LLF partage avec l'UR CLILLAC-ARP une plateforme de linguistique expérimentale sous la responsabilité d'une ingénieure d'étude de l'unité qui gérait durant la période d'évaluation une salle d'expérience avec deux cabines acoustiques. Le LLF a fortement investi dans l'équipement de la plateforme, composée actuellement de deux oculomètres fixes, un système de lunettes oculométriques, un système d'électro-encéphalographie, un système de captation de la conductivité de la peau et un échographe. Le LLF administre et utilise également une plateforme de linguistique informatique, qui comporte plusieurs serveurs de calcul et de stockage, un serveur dédié aux expériences psycholinguistiques en ligne et à la démonstration des ressources développées au laboratoire, le tout représentant un investissement de plus de 80 k€. L'activité de calcul de l'unité s'appuie également sur le supercalculateur Jean Zay du CNRS pour ses travaux les plus ambitieux, comme l'entraînement de grands modèles de langue.

## Points faibles et risques liés au contexte pour les quatre références ci-dessus

Certains chercheurs restent moins identifiés au-delà de la communauté scientifique nationale que d'autres. L'unité est encouragée à mettre en place des dispositifs favorisant la visibilité et le rayonnement scientifique de tous ses membres, par exemple en proposant des aides à la traduction, une liste de revues ciblées comme prioritaires dans différents domaines, en promouvant la co-publication, la systématisation des profils ORCID, HAL et des différents types d'accès libre pour les publications.

L'unité a une politique restrictive liée au potentiel de financement des futurs docteurs, ce qui amène à un nombre limité de candidats/doctorants chaque année. À titre d'exemple, le nombre de personnes inscrites fin 2022 est de 29, en baisse de 25 % par rapport au début du contrat. Cette importante baisse de jeunes chercheurs pourrait être analysée et un correctif pourrait être apporté.

L'unité a une stratégie de communication assez hétérogène. Les séminaires, formations, nouvelles productions de l'unité circulent principalement dans les listes de diffusion du laboratoire mais rarement publiquement. L'unité est encouragée à investir plus dans la communication et la diffusion de ses recherches, événements organisés, formations, etc., par exemple, en mettant régulièrement à jour son site, en utilisant les listes de diffusion nationales et internationales, ou en se servant des réseaux GDR dans lesquels elle est impliquée.

En matière de locaux, l'unité est tributaire de la politique de la tutelle hébergeuse, l'université Paris Cité. Elle dispose actuellement d'une surface de bureaux importante de 580 m<sup>2</sup> mais que les règles de sécurité ne permettent pas d'utiliser de manière optimale. Les 67 postes de travail dont elle dispose, pour un effectif moyen de 70 membres auxquels s'ajoutent neuf émérites, ainsi que chaque année une dizaine de visiteurs et une trentaine de stagiaires, rendent l'accueil des chercheurs relativement tendu.

En ce qui concerne l'infrastructure scientifique et les ressources documentaires, le laboratoire consacre une part importante de son budget à l'équipement de ses deux plateaux techniques : la plateforme de linguistique expérimentale, partagée avec l'UR CLILLAC-ARP, et la plateforme de linguistique informatique. Environ 150 k€ ont été dégagés sur la période d'évaluation pour leur équipement, principalement grâce à la mobilisation des ressources propres, mais aussi par des financements spécifiques des tutelles, la dotation d'État venant en soutien à la finalisation de certains achats. Certains de ces matériels acquis dans les dernières années, comme le système EEG, ont tardé à être utilisés en raison de la crise sanitaire, mais ils ne fonctionnent pas aujourd'hui à leur plein potentiel.

Les deux plateformes jouent un rôle important dans la croissance de l'activité du laboratoire en linguistique expérimentale et en linguistique computationnelle. Toutefois, leur développement est freiné par un manque de moyens humains qui ne permet pas une utilisation optimale de ces ressources. La plateforme de linguistique expérimentale était gérée par une unique IE qui y consacrait 50 % de son activité. La plateforme de linguistique informatique s'appuie sur la collaboration de trois ingénieurs, mais cela semble insuffisant pour répondre à l'ensemble des demandes.

## DOMAINE 3 : PRODUCTION SCIENTIFIQUE

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

La production scientifique du laboratoire est remarquable par sa quantité et sa qualité, sur des supports de publication internationaux et reconnus dans leurs domaines. La politique du laboratoire incite fortement les personnels d'appui à la recherche à publier, ainsi que les étudiants dès le premier niveau (master) et les post-doctorants. Les effets bénéfiques de cette politique sont visibles dans la production et le nombre de publications signées en collaboration. L'équilibre entre les thématiques de recherche est également très satisfaisant. L'appréciation globale est très positive.

- 1/ La production scientifique de l'unité satisfait à des critères de qualité.*
- 2/ La production scientifique de l'unité est proportionnée à son potentiel de recherche et correctement répartie entre ses personnels.*
- 3/ La production scientifique de l'unité respecte les principes de l'intégrité scientifique, de l'éthique et de la science ouverte. Elle est conforme aux directives applicables dans ce domaine.*

Points forts et possibilités liées au contexte pour les trois références ci-dessus

La production scientifique du laboratoire est remarquable par sa qualité, sa quantité et son équilibre. L'unité a pris en compte les recommandations de la précédente évaluation en augmentant la part de sa production dans des revues. La part des contributions dans des actes de colloques reste élevée et correspond également à des supports de publication reconnus mondialement, les actes de colloques étant la référence pour certains champs disciplinaires, en particulier dans les domaines du traitement automatique des langues et de la phonétique et phonologie : LREC (Language Resources and Evaluation Conference), COLING (International Conference on Computational Linguistics), ACL (Association of Computational Linguistics), EMNLP (Conference on Empirical Methods in Natural Language Processing, Interspeech, SpeechProsody, etc. Un effort collectif a été fait pour favoriser les publications dans les supports de qualité et de haute visibilité. Les résultats sont visibles dans l'examen des supports de publications dont la qualité est reconnue internationalement : Language, Journal of linguistics, Morphology.

Le comité signale également comme particulièrement remarquable la parution de la GGF (Grande Grammaire du Français) sur la période d'évaluation, fruit d'un travail de près de vingt années, ayant mobilisé plusieurs membres du laboratoire et dont la direction scientifique a été assurée par des membres du laboratoire. Un important travail de communication et diffusion auprès du public a en outre été fait autour de cette parution, ce qui est tout à fait exceptionnel.

La production scientifique montre une grande cohérence et une répartition équilibrée entre les domaines thématiques et les axes transversaux, le laboratoire ayant opté pour une organisation en thèmes et axes méthodologiques plutôt qu'en équipes. Cette organisation permet une forte collaboration entre les différents pôles et renforce la cohérence du laboratoire, comme le montrent les productions en collaboration entre plusieurs membres, dont la répartition varie en fonction des thèmes ou des méthodes. Le laboratoire mène également une politique d'incitation à la publication auprès des doctorants (et même des étudiants de master, fait beaucoup plus rare) et des personnels d'appui à la recherche, politique dont les retombées positives sont visibles dans la production scientifique, à travers plusieurs publications collectives (sur 208 articles de revues, sept d'entre eux sont issus d'un travail de Master, 62 ont reçu la signature d'un doctorant et 37 d'un post-doctorant, comme premier auteur dans 63 % des cas). Cette politique de travail en collaboration contribue certainement à l'incitation à la publication dans des supports de la plus haute qualité, instaurant ainsi un cercle vertueux.

Le laboratoire veille également à l'éthique dans ses projets, avec une incitation à passer par le comité d'éthique de la tutelle de l'université Paris Cité, dont deux membres du laboratoire font partie, pour la validation des projets. Le laboratoire vise à étendre cette pratique, obligatoire dans certaines disciplines, à l'ensemble des projets. Dans le cadre du Master PGSL, il a également été mis en place un comité pédagogique spécifique d'examen des projets de Master qui est pour moitié composé de membres du laboratoire. Cette initiative est tout à fait remarquable.

Le laboratoire a enfin mis en place une politique de science ouverte, tout en respectant les modes de publication traditionnels pour les disciplines dont c'est encore le principal moyen de visibilité et reconnaissance. Les actions mises en place sont notamment le refus de subventionner directement les publications qui ne seraient pas en libre accès, le refus de frais prohibitifs de publication, et un soutien régulier (par le biais de subventions récurrentes et l'investissement de membres du laboratoire dans les comités éditoriaux ou de direction) des projets éditoriaux de sciences ouvertes, comme l'éditeur Language Science Press et la revue *Glossa*.

## Points faibles et risques liés au contexte pour les trois références ci-dessus

Le laboratoire compte quelques personnalités motrices qui influent sur les publications tant en termes de qualité qu'en termes de quantité. Pour les articles de revues, quatre auteurs produisent à eux seuls 28 % des publications ; pour les chapitres d'ouvrages, trois auteurs totalisent 45 % des publications (dans des contributions toutefois toujours co-signées). Les moyennes de taux de publications annuel par personnel de recherche permanent se situent entre 2,7 (moyenne) et 2,3 (médiane). Elles sont tirées vers le haut par les auteurs moteurs, dont le taux de publication peut aller jusqu'à dix-sept contributions par an. Ce taux est également tiré vers le haut par la parution de la *GGF*, collectif dont les chapitres sont comptabilisés comme chapitres d'ouvrage. Les membres du laboratoire y totalisent 41 chapitres sur 148 (soit plus du quart). La production scientifique risque néanmoins de diminuer lors du prochain bilan en fonction des états d'aboutissement des différents projets ou des possibles mutations ou départ de ces membres les plus productifs.

La production du laboratoire, bien que comportant des publications phares et dans des supports de qualité, comporte très peu de monographies et de directions d'ouvrages (respectivement huit et neuf sur la période). Ce chiffre est stable par rapport au précédent rapport, mais reste bas pour un laboratoire qui a une activité globale aussi dynamique. Les monographies et ouvrages sont aussi importants pour faire la somme des connaissances actuelles dans une discipline.

Le DAE donne à voir une grande cohérence affichée dans les différents axes disciplinaires et transversaux, mais il n'est pas précisé par quel biais cette cohérence est assurée. Là encore, l'évolution risque d'être fortement fonction des personnalités motrices et des projets phares, au risque de laisser dans l'ombre des recherches de personnels plus isolés ou hors projet financé.

À propos de la publication des données de la recherche, le rapport ne permet pas de mesurer la politique mise en place par le laboratoire et son effet : il n'est fourni aucune indication précise sur les publications de codes sources, ressources et corpus. Le comité attire l'attention du laboratoire sur le fait que dans une politique de science ouverte, la publication des données de la recherche (codes sources, ressources, corpus, données) est tout aussi centrale que la publication de contributions scientifiques résultant de l'analyse de ces données. Sur cette question, le laboratoire mentionne en outre préférer des sites de dépôts internationaux, plus adaptés à leur public, ce qui est certes important, mais comporte le risque d'être dépendant de ces dépôts. Le laboratoire, en tant qu'acteur central dans la recherche linguistique en France, est peu impliqué dans les politiques de publication des données de la recherche, en termes d'animation des réseaux ou des centres de dépôts nationaux et autres infrastructures. Bien qu'affichant une politique de science ouverte, le DAE montre que le laboratoire porte peu de projets de diffusion ouverte de la recherche, en dehors de deux soutiens à des initiatives extérieures.

Pour ce qui est des produits de la recherche et toujours dans une politique de science ouverte, le DAE fait assez peu mention de la prise en compte des principes FAIR (Findable, Accessible, Interoperable, Reusable) dans ses projets et d'une éventuelle politique de laboratoire sur ce point.

D'une manière générale, le DAE ne permet pas de mesurer précisément l'implication de ses membres dans les activités de diffusion et valorisation de la recherche en tant que porteurs et non utilisateurs : direction de revues ou de collection, implication dans des comités d'édition ou de rédaction. De même, la production scientifique ne fait état que de deux recensions sur la période. L'activité de recension peut certes être considérée comme de second ordre dans la production scientifique, mais elle contribue également à la diffusion de la recherche.

En ce qui concerne l'éthique, le DAE ne mentionne pas la prise en compte du plan de gestion des données, dont l'utilisation doit pourtant être développée lors de la soumission de projets, et de la formation des personnels à ce sujet.

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

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

Le LLF affiche un niveau d'interactions tout à fait satisfaisant avec le monde non académique et fait preuve d'une capacité à mener à bien des projets ambitieux impliquant une variété d'acteurs, comme le projet européen SIGNHUB (dans le cadre du programme Horizons 2020). L'unité interagit avec le secteur économique, éducatif et médical, collabore avec différentes communautés de locuteurs et mène des actions de vulgarisation grand public : podcast « Parler comme jamais », communications autour de la GGF dans des émissions de radio et la presse écrite, etc. Ces interactions s'articulent avec les activités de recherche fondamentale du laboratoire. Il faut également saluer l'effort de diffusion des résultats des recherches menées à travers la mise à disposition des ressources développées.

- 1/ L'unité se distingue par la qualité et la quantité de ses interactions avec le monde non académique.*
- 2/ L'unité développe des produits à destination du monde culturel, économique et social.*
- 3/ L'unité partage ses connaissances avec le grand public et intervient dans des débats de société.*

### Points forts et possibilités liées au contexte pour les trois références ci-dessus

Les recherches menées dans un des axes transversaux du laboratoire, la « linguistique descriptive et de terrain », sont le lieu d'interactions avec différentes communautés de locuteurs ainsi qu'avec le secteur éducatif. Elles donnent lieu à des actions de communication vers le grand public. Elles ont conduit à des développements aux retombées sociétales directes. En linguistique descriptive, la GGF, synthèse monumentale et unique faisant une large place à la description de la langue orale, constitue une opportunité d'interaction avec les enseignants du secondaire, sous la forme de formations dispensées par une des éditrices de l'ouvrage. Ce même champ de la linguistique descriptive a contribué à la production de ressources pour les langues parlées dans l'espace francophone par le site d'un projet institutionnel, « Langues et grammaires du monde dans l'espace francophone ».

Les recherches menées en linguistique de terrain, avec les populations concernées — Sourds d'Europe et autochtones —, notamment en recourant à la science participative, permettent d'articuler description linguistique et préservation du patrimoine historique et culturel. Le projet SIGNHUB, mené dans le cadre du programme H-2020, a débouché sur la conception d'une base de connaissances sur les langues des signes, mais également sur la vie des malentendants d'Europe, base qui a été mise à disposition en ligne.

Les activités de linguistique descriptive et de linguistique de terrain s'accompagnent d'actions de vulgarisation à destination du grand public, soit directement à travers la participation à des podcasts, à un film documentaire (We were there... we were here), soit par plusieurs médias à large audience, comme des émissions radiophoniques ou des articles dans la presse écrite nationale et internationale (e.g. le New York Times). Outre la meilleure connaissance de la recherche en linguistique par le grand public qu'assurent les actions de vulgarisation, celles-ci s'insèrent pleinement dans les débats de société et y contribuent concrètement.

Une retombée sociétale claire que la communication menée autour des travaux en linguistique descriptive et en linguistique de terrain pourrait générer serait de favoriser une plus grande inclusivité à travers, d'une part, la visibilité des langues minorées et donc de celle de leurs communautés de locuteurs et, d'autre part, la déconstruction d'idées reçues et la dé-stigmatisation de certaines pratiques langagières et, partant, de certains locuteurs.

Les interactions avec le monde proprement économique émanent essentiellement de l'axe linguistique informatique et se concrétisent par le financement de trois thèses sur dispositif Cifre avec des cabinets de conseil (Quantmetry, OnePoint et IPSIDE) et par celui de recherches en synthèse de la parole.

Il convient enfin de saluer la diffusion par le LLF des ressources qu'il développe, qu'il s'agisse de logiciels, de corpus ou de bases de données, mises à disposition gratuitement et sous licences libres pour la communauté académique et, pour certaines, payantes pour un usage commercial.

## Points faibles et risques liés au contexte pour les trois références ci-dessus

Deux axes transversaux du contrat 2019-2023, qui ont vocation à se maintenir comme axes de recherche dans le futur contrat, sont absents du volet « interaction avec le monde non académique » dans le DAE. Le comité conçoit assez bien que l'axe « formalismes linguistiques » se prête moins directement que les autres à des interactions avec le monde non académique, mais il s'étonne qu'aucune mention ne soit faite pour l'axe « linguistique expérimentale ».

La visite a permis de prendre connaissance de la participation de cet axe à la fête de la science et l'intention d'intervenir dans les écoles et les crèches. Cette dernière activité n'entre pas, cependant, dans le cadre du présent contrat.

Les interactions avec le domaine privé, qui émanent principalement de l'axe « linguistique computationnelle » sont des sources importantes de financement de thèses dans le cadre du dispositif Cifre et, dans le présent contrat, de recherches en synthèse de la parole en partenariat avec Google. Si la recherche de financements s'insère dans le programme scientifique préétabli, sans contraindre ni le choix des thématiques étudiées, ni la diffusion des recherches menées et des produits qui en découlent, alors le bénéfice pour les sociétés privées et la recherche académique est bien mutuel. Le comité craint toutefois que le choix de certaines thématiques de recherche a été et soit en partie guidé par l'opportunité de financements privés. Le risque de perte de cohérence scientifique et d'indépendance est alors réel.

Le bénéfice pour LLF de travailler avec un cabinet de conseil en pédagogie plutôt que directement avec les enseignants du primaire, secondaire ou supérieur, avec lesquels l'unité collabore déjà, n'apparaît pas clairement. Pour ce qui est des formations dispensées auprès des enseignants du secondaire autour de la GGF, il a été mentionné lors de la visite qu'elles sont dispensées par une seule enseignante-chercheuse et non par les éditrices, comme le souligne le DAE.

## ANALYSE DE LA TRAJECTOIRE DE L'UNITÉ

La nouvelle structuration proposée vise à satisfaire l'objectif ambitieux défini dans le précédent projet : jouer un rôle moteur au plan mondial dans le tournant empirique de la linguistique, en combinant une montée en puissance de la conception et de l'exploitation de ressources linguistiques numériques (corpus et bases de données), et une généralisation des méthodes relevant de la linguistique expérimentale. Le renforcement de la thématique linguistique computationnelle a permis de viser d'une part la production d'algorithmes et de systèmes de traitement automatique des langues, et d'autre part une meilleure intégration de techniques computationnelles pour l'étude de la langue. Un bilan du dernier contrat à l'aune de ces objectifs, conduit à souligner une réelle montée en compétence en linguistique empirique. LLF est devenu une référence en France pour la linguistique empirique. En ce qui concerne la linguistique computationnelle, l'intégration de cette thématique a bien commencé mais elle n'est certainement pas encore optimale. C'est clairement une opportunité que la trajectoire de l'unité doit saisir. Il est en effet assez rare d'avoir des compétences en intelligence artificielle et en apprentissage automatique dans un laboratoire de linguistique générale.

Au cours du présent contrat, le LLF a bien renforcé son pôle en phonétique et phonologie de laboratoire. Il est devenu crucial que le dispositif de structuration de l'unité fasse toute sa place à cette activité. La nouvelle structuration en axes a permis une répartition plus équilibrée des membres, quels que soient leurs objets d'étude, sachant qu'un même membre peut toujours appartenir à plusieurs axes. Ce choix sera maintenu et davantage concrétisé par des réunions périodiques (au minimum annuel, et plus selon la dynamique de chaque axe) qui permettront aux membres du LLF de bénéficier mutuellement de la grande richesse de compétences du laboratoire.

Le séminaire hebdomadaire, le LingLunch, a été un succès qui attire des chercheurs nationaux et internationaux de très haut niveau aussi bien comme intervenant que comme auditeur. Il continuera bien sûr dans le prochain contrat à témoigner du rayonnement du laboratoire. En ce qui concerne les formalismes linguistiques, LLF est un centre internationalement reconnu pour le développement d'analyses de pointe dans une variété de formalismes grammaticaux, et notamment le modèle minimaliste et HPSG en syntaxe, la sémantique formelle, la sémantique distributionnelle, la morphologie gabaritique, la morphologie réalisationnelle, la nanosyntaxe. Toutes ces activités seront maintenues et développées dans le prochain contrat.

L'unité continuera aussi à se servir de son expertise actuelle en matière d'utilisation de diverses méthodes (production, perception, compréhension ; méthodes en temps réel et offline), en renforçant les méthodes expérimentales sur internet, une des forces de l'équipe. L'obtention du robot FurHat, qui possède des capacités multimodales de pointe, permettra enfin à l'équipe d'examiner des phénomènes multimodaux tels que le rire et les gestes faciaux, dans une diversité de langues parlées (langues de Chine, créoles français), afin de proposer des généralisations dans une perspective interlangue, tout en permettant d'explorer la diversité linguistique.

Des synergies avec l'axe linguistique computationnelle sont également attendues pour la mise en place d'outils pouvant faciliter la transcription de données dans ce type de langues. Grâce à ces synergies, qui ont émergé pendant les dernières années, l'axe corpus, description et terrain devrait avoir un périmètre plus large.

## RECOMMANDATIONS À L'UNITÉ

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

Le LLF est encouragé à inciter ses personnels à suivre des formations en langue française et en langue anglaise sur leur temps de travail, afin de permettre les échanges fluides dans l'une et l'autre langue.

L'unité est incitée à faire valoir auprès de ses tutelles ses besoins en espaces de travail supplémentaires, afin notamment d'être à même de poursuivre sa politique d'accueil de visiteurs internationaux et de stagiaires en master.

Le comité recommande à l'unité d'adopter rapidement son règlement intérieur et l'invite à mettre en place une politique en matière de développement durable, dans la limite de son champ de compétences.

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

L'unité est encouragée à investir dans la communication et la diffusion de ses recherches, formations et événements organisés. Cela peut se faire en maintenant régulièrement à jour son site web, en utilisant les listes de diffusion nationales et internationales, et en exploitant les réseaux GDR auxquels elle participe.

De plus, elle est encouragée à instaurer des dispositifs visant à accroître la visibilité et le rayonnement scientifique de l'ensemble de ses membres. Par exemple, elle pourrait proposer des aides à la traduction, établir une liste de revues prioritaires, promouvoir la co-publication, encourager la mise à jour et la généralisation des profils ORCID, HAL, ainsi que de promouvoir les divers modes d'accès libre aux publications de ses membres.

Étant donné que certains équipements de la plateforme expérimentale ne sont pas pleinement exploités ou fonctionnent en deçà de leur pleine capacité, il est suggéré de réviser l'affectation des ingénieurs dédiés à ces plateformes, de promouvoir la réservation en ligne, et de veiller à une exploitation optimale des dispositifs.

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

L'unité est encouragée à poursuivre sa politique actuelle d'incitation à la publication pour les personnels d'appui à la recherche, les étudiants et les post-doctorants, dont les retombées positives sont déjà visibles. La politique d'augmentation des publications dans des supports de qualité, et la mise en place d'une politique d'éthique de la recherche doit être étendue à l'ensemble du laboratoire.

Le comité recommande également à l'unité de poursuivre et de renforcer sa politique de science ouverte en augmentant le soutien à des projets, en développant la formation sur l'ouverture de la science auprès des jeunes chercheurs, et en adoptant une politique de bonnes pratiques pour la pérennisation des données. L'unité est ainsi encouragée à renforcer la pérennité des données et des ressources qu'elle produit par une mise à disposition sur des dépôts identifiés nationaux ou internationaux (Huma-Num, Ortolang, CLARIN) dont la gestion ne repose plus sur les personnels d'appui de la recherche de l'unité.

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

L'unité est encouragée à poursuivre l'effort de diffusion de ses recherches auprès du grand public et dans le secteur éducatif. Le travail de formation des enseignants du secondaire autour de la GGF pourrait impliquer davantage de membres de l'unité.

L'unité est également incitée à continuer son travail en collaboration avec les communautés de locuteurs.

En matière de partenaires du secteur économique, il est nécessaire de veiller à renforcer une réelle collaboration scientifique et à limiter les simples services.

Le comité recommande enfin que l'axe « linguistique expérimentale » s'implique plus nettement dans les interactions non académiques.

## DÉROULEMENT DES ENTRETIENS

### DATE

**Début :** 27 novembre 2023 à 09h30

**Fin :** 27 novembre 2023 à 17h00

**Entretiens réalisés : en distanciel**

### PROGRAMME DES ENTRETIENS

9h45- 10h	
10h-10h30	Réunion à huis-clos avec les tutelles Mme Marie Salaün, Vice-doyenne recherche de la faculté Sociétés et Humanités, université Paris Cité Mr Ricardo Extepare, Directeur Adjoint Scientifique, INSHS, CNRS Mme Marie-Hélène Papillon Déléguée régionale CNRS
10h30h-11h30	Session plénière Présentation du laboratoire : Olivier Bonami, directeur de l'unité ; Giuseppina Turco, directrice adjointe; Heather Burnett, future directrice
11h30-11h45	Pause
11h45-12h15	Réunion (sans la direction de l'unité) <u>Domaine thématique 1 : Le mot</u> Introduction : Berthold Crysmann, responsable du domaine
12h15-12h45	Réunion (sans la direction de l'unité) <u>Domaine thématique 2 : La phrase</u> Introduction : Anne Abeillé, responsable du domaine
12h45-13h15	Réunion (sans la direction de l'unité) <u>Domaine thématique 3 : Discours et dialogue</u> Introduction : Jonathan Ginzburg, responsable du domaine
13h15-14h	Déjeuner
14h-14h30	Réunion (sans la direction de l'unité) <u>Domaine thématique 4 : Variation et changement</u> Introduction : Heather Burnett, responsable du domaine
14h30-15h	<u>Les axes transversaux</u> Introduction : Benoît Crabbé, responsable de l'axe « linguistique computationnelle »
15h-15h15	Pause
15h15-15h45	Réunion à huis-clos avec les doctorants Introduction : Marie Legentil, doctorante en 3e année
15h45-16h15	Réunion à huis-clos avec les post-doctorants et les visiteurs Introduction : Lucie Janků, postdoctorante
16h15-16h45	Réunion à huis-clos avec les ingénieurs, techniciens et administratifs Ali Dhibi, Sophie Lempérière, Achille Falaise, Loïc Liégeois, Alexandre Roulois Introduction: Loïc Liégeois, représentant des ITA au conseil de laboratoire
16h45-17h15	Réunion à huis-clos avec la direction de l'unité (actuelle et future) Olivier Bonami (directeur actuel), Giuseppina Turco (directrice adjointe actuelle), Heather Burnett (future directrice), Marie-Hélène Candito (future directrice adjointe)
17h15-17h45	Réunion à huis clos du comité d'experts en présence du conseiller scientifique`
17h45	Fin de la visite

## OBSERVATIONS GÉNÉRALES DES TUTELLES

Le Président

Paris, le 14 février 2024

HCERES  
2 rue Albert Einstein  
75013 Paris

**Objet : Rapport d'évaluation de l'unité DER-PUR250024342 - LLF - Laboratoire de linguistique formelle.**

Madame, Monsieur

L'Université Paris Cité (UPCité) a pris connaissance du rapport d'évaluation de l'Unité de Recherche **LLF - Laboratoire de linguistique formelle.**

Ce rapport a été lu avec attention par la direction de l'unité qui signale des erreurs factuelles à corriger (cf courrier joint), la vice-doyenne Recherche et le doyen de la Faculté Sociétés & Humanités d'UPCité (cf courrier joint), par la vice-présidente Recherche d'UPCité et par moi-même.

**Présidence**

**Référence**

Pr/DGDRIVE/2023

**Affaire suivie par**

Christine Debydeal -  
DGDRIVE

**Adresse**

85 boulevard St-Germain  
75006 - Paris

J'adresse mes remerciements au comité, et vous informe ne pas avoir d'observations de portée générale à apporter.

Je vous prie d'agréer, Madame, Monsieur, l'expression de ma considération distinguée.

[www.u-paris.fr](http://www.u-paris.fr)

Édouard Kaminski



Monsieur Sylvain Moutier  
Doyen de la Faculté Sociétés & Humanités (SH)  
Université Paris Cité  
2 rue Valette  
75005 Paris

*Paris, le 12 février 2024*

**Objet : Retour de la Faculté SH sur le rapport provisoire d'évaluation HCERES du  
Laboratoire DER-PUR250024342 - LLF - Laboratoire de linguistique formelle.**

Chères et chers Collègues,

Concernant le rapport provisoire d'évaluation HCERES de cette unité (LLF), la Faculté SH n'a aucune observation à signaler (ni générale, ni factuelle), au-delà des remarques factuelles très détaillées déjà communiquées par la direction de ce laboratoire.

En vous remerciant de votre attention et en vous priant, chères et chers Collègues, d'accepter mes chaleureuses salutations,



Sylvain Moutier  
Doyen de la Faculté Sociétés & Humanités

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)