

RAPPORT D'ÉVALUATION DE L'UNITÉ

CLILLAC-ARP - Centre de Linguistique
Interlangue, de Lexicologie, de Linguistique
Anglaise et de Corpus, Atelier de Recherche sur
la Parole

SOUS TUTELLE DES ÉTABLISSEMENTS ET ORGANISMES :

Université Paris Cité

CAMPAGNE D'ÉVALUATION 2023-2024
VAGUE D

Rapport publié le 05/04/2024



Au nom du comité d'experts :

Jean-Marie Klinkenberg, 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

Président :	M. Jean-Marie Klinkenberg, Université de Liège, Belgique
	M. Gilles Adda, CNRS, Orsay (personnel d'appui à la recherche)
	M. Jean-Michel Fournier, Université de Tours
Expert(e)s :	M. Laurent Gautier, Université de Bourgogne
	M. Alain Peyraube, École des Hautes Études en Sciences Sociales, Paris
	Mme Céline Poudat, Université Côte d'Azur (représentante du CNU)

REPRÉSENTANT DU HCÉRES

M. Bernard Laks

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

M. Mathieu Arnoux, Université Paris Cité
Mme Aouatif El Fakir, Université Paris Cité

CARACTÉRISATION DE L'UNITÉ

- Nom : Centre de Linguistique Interlangue, de Lexicologie, de Linguistique Anglaise et de Corpus, Atelier de Recherche sur la Parole
- Acronyme : CLILLAC-ARP
- Label et numéro : 3967
- Composition de l'équipe de direction : Mme Natalie Kübler (Directrice) et Mme Agnès Celle (Directrice adjointe)

PANELS SCIENTIFIQUES DE L'UNITÉ

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

THÉMATIQUES DE L'UNITÉ

Les thématiques de l'unité sont structurées en deux axes, subdivisés en cinq sous-axes :

— Axe 1 : Sémantique, Discours et Parole. Les sous-axes sont :

- sa1 : Parole et Diversité, phonétique et phonologie ;
- sa2 : Syntaxe et Sémantique.

Ces sous-axes sont articulés par deux axes transversaux : Discours et Société, et Comparaison, Contraste, Variation.

— Axe 2 : Langues et Discours de Spécialité, Traductologie, Interculturalité. Les sous-axes sont ici :

- sa1 : Discours spécialisés : entre phraséologie et terminologie ;
- sa2 : Traduction pragmatique et communication spécialisée ;
- sa3 : Politiques linguistiques et interculturalité.

Ceux-ci sont reliés par une transversale : Linguistique de corpus : variation, normes, genres textuels.

Les thématiques correspondant à cette structure peuvent être détaillées comme suit :

— A1sa1 : socio-phonétique et variétés de l'anglais, avec une forte composante expérimentale (projet Phonologie de l'Anglais Contemporain ; variation de voix et accents, expertise vocale en criminologie et sa représentation, perspective contrastive sur la syllabe ; relation entre variation et acquisition, et entre variation et diachronie, à partir de grands corpus de parole ; discours d'apprenants et acquisition L2, à visée pédagogique mais aussi phonologie de l'interlangue).

— A1sa2 : interfaces et constructions (ellipse et anaphore ; catégorie du nom, détermination nominale et interfaces nom/discours ; concept de constructions/multimodalité).

— Discours et société (médias sociaux ; expression des émotions et emphase dans le discours politique et militant).

— Études contrastives (phrase averbale, questions non canoniques et marqueurs discursifs ; système modal de l'anglais : variation sociolinguistique et diachronique).

— A2sa1 : discours spécialisés (constructions polylexicales, expressions idiomatiques et parémies ; collocations génériques ; groupes nominaux complexes, variation terminologique, néologie, nominalisation et métaphore grammaticale ; caractérisation de discours idéologiques et argumentatifs).

— A2sa2 : articulation recherche/formation/sphère professionnelle (révision et post-édition, en particulier dans le contexte de la traduction automatique ; apports de la phraséologie et la terminologie à la traduction spécialisée).

— A2sa3 : interprétation en service public et médiation linguistique ; développement et rôle des solutions digitales dans l'intégration des migrants ; impact cognitif de la gestualité co-verbale ; conditions d'émergence du développement linguistique, linguistique du développement.

Il y a enfin un « axe transversal à l'unité » : Politique de la donnée et interopérabilité des corpus écrits et oraux. Il articule ambitieusement l'ensemble des sous-axes dans les domaines de la génération de corpus, de l'épistémologie de la science des données linguistiques et de la conceptualisation des jeux de données, complétée de déclinaisons expérimentales et pédagogiques.

HISTORIQUE ET LOCALISATION GÉOGRAPHIQUE DE L'UNITÉ

Le CLILLAC-ARP (Centre de Linguistique Interlangue, de Lexicologie, de Linguistique Anglaise et de Corpus - Atelier de Recherche sur la Parole) est une unité née de deux fusions successives. En 2006, le CIEL (Centre Interlangue d'Études de Lexicologie) de l'Unité de Formation et de Recherche (UFR) EILA (Études interculturelles de langues appliquées) et l'équipe LILA (Linguistique Interlangue et Linguistique Anglaise), de l'UFR d'Études Anglophones, ont fusionné au sein du CLILLAC. Peu de temps après, en 2008 (soit avant la fusion des universités Paris-Descartes et Paris-Diderot en 2020), l'ARP de l'UFR de Linguistique est venue rejoindre le CLILLAC. Au terme de ces mutations, le CLILLAC-ARP constitue une équipe de 64 membres dont les EC (enseignants-chercheurs)

— au nombre de 38 — relèvent des trois UFR de l'université Paris Cité mentionnées.

Initialement, l'unité organisait son travail le long de trois axes thématiques, reliés par des axes transversaux : tandis que les collègues de linguistique appliquée continuaient de former un axe spécifique, les autres se répartissaient entre un axe Syntaxe, Sémantique et Pragmatique, et un axe Discours, Parole et Cognition réunissant l'ensemble des membres s'intéressant à l'oral et ses manifestations. En 2018, après un premier contrat consacré à consolider l'unité et à établir sa cohésion, et suivant en cela les recommandations de la précédente évaluation, une nouvelle répartition était décidée. Le nombre d'axes a été réduit à deux (deux des axes antérieurs fusionnant) : un axe « Sémantique, Discours, Parole », qui rassemble notamment des phonéticiens et des phonologues de l'UFR de Linguistique et de l'UFR d'études anglophones, mais aussi des spécialistes de syntaxe, de sémantique et de pragmatique, et un axe « Langues et Discours de Spécialité, Traductologie, Interculturalité », qui associe tous les membres de l'UFR EILA. Mais cette réduction a été compensée par le maintien d'une subdivision en sous-axes et une multiplication des axes transversaux. L'axe 1 se subdivise ainsi désormais en deux sous-axes décrits en détail à la rubrique « Thématiques de l'unité », sous-axes reliés par deux axes transversaux. L'axe 2 se divise quant à lui en trois, un axe transversal venant les structurer. Enfin, un axe transversal commun à l'ensemble de l'unité — axe qui existait déjà dans la configuration précédente — vient compléter cette architecture : « Politique de la donnée et interopérabilité des corpus écrits et oraux ».

L'unité est rattachée à l'École Doctorale 622 Sciences du Langage, co-accréditée avec Sorbonne Nouvelle.

Depuis 2013, tous les membres de l'unité, qui étaient jusque-là dispersés sur trois sites, jouissent d'espaces de travail répartis dans un seul édifice : le bâtiment Olympe de Gouges, sur le site de Paris Rive Gauche, Place Paul Ricœur, 75013 Paris. Ils n'y disposent toutefois d'aucun espace de convivialité.

ENVIRONNEMENT DE RECHERCHE DE L'UNITÉ

L'unité de recherche CLILLAC-ARP affiche très clairement un positionnement à la fois raisonné et ambitieux dans l'environnement de recherche disciplinaire qui est le sien.

Il se manifeste tout d'abord au niveau local et régional, à travers des interactions avec plusieurs unités de premier plan par des accueils d'enseignants-chercheurs en délégation en provenance de l'UMR 7110 CNRS – Paris Cité « Laboratoire de Linguistique Formelle » (LLF) ou de l'UMR 7018 CNRS–Sorbonne Nouvelle « Laboratoire de Phonétique et de Phonologie » (LPP). Des interactions sont également mentionnées dans le document d'autoévaluation (ci-après DAE) avec l'UMR CNRS - Paris-Saclay - INRIA - CentraleSupélec « Laboratoire Interdisciplinaire des Sciences du Numérique » (LISN, anciennement LIMSI) et avec l'UMR CNRS-Sorbonne Université « Institut des Systèmes Intelligents et Robotique » (ISIR), mais leur nature n'est pas précisée.

Au niveau national ensuite, l'unité entretient des collaborations institutionnalisées avec un grand nombre d'équipes et d'institutions à travers la France : avec le Service National de Police Scientifique (Écully), l'Institut de Recherche Criminelle de la Gendarmerie Nationale, l'université d'Avignon, Aix-Marseille Université ou encore avec l'université Toulouse Jean-Jaurès.

Au niveau européen et international enfin, le laboratoire mise beaucoup sur la place de Paris Cité dans l'Alliance Universitaire européenne Circle U. Même si tous les dispositifs ne sont pas encore en place, plusieurs projets ont déjà pu être développés avec des universités désormais membres de l'Alliance (par exemple Deep Learning for Language Assessment (DLLA) ou Multilingual Student Transaltion (MUST), etc.). Un point d'attention mérite toutefois d'être mentionné : si les corpus sur lesquels l'unité a travaillé jusqu'à présent entrent en effet en résonance avec les domaines thématiques retenus par l'Alliance (santé, sciences de la terre), il faudra veiller à ce que les sciences du langage ne restent pas une discipline auxiliaire par rapport aux disciplines expérimentales. Pour le reste, les contrats formels signés avec des établissements d'enseignement supérieurs européens, et a fortiori extra-européens, sont de taille modeste pour une unité d'une telle importance. Il reste là une indéniable marge de progression.

L'unité affiche par ailleurs son ambition en étant bien intégrée, au niveau de l'axe 1, au Labex EFL (« Empirical Foundations of Linguistics »), avec des retombées tant pour les enseignants-chercheurs que pour les doctorants ou les équipements. L'ambition affichée d'inclure l'intégralité de l'unité au moment du renouvellement du Labex ne peut être que saluée et devra bénéficier du soutien de l'établissement.

Enfin, les Masters adossés à l'unité sont intégrés à l'EUR PGSL (Paris Graduate School of Linguistics), qui doit permettre d'asseoir le continuum master-doctorat et de déployer les recrutements à l'international. C'est un point fort de l'unité.

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

Catégories de personnel	Effectifs
Professeurs et assimilés	10
Maîtres de conférences et assimilés	28
Directeurs de recherche et assimilés	0
Chargés de recherche et assimilés	0
Personnels d'appui à la recherche	2
Sous-total personnels permanents en activité	40
Enseignants-chercheurs et chercheurs non permanents et assimilés	1
Personnels d'appui non permanents	0
Post-doctorants	0
Doctorants	23
Sous-total personnels non permanents en activité	24
Total personnels	64

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	38	0	2
Total personnels	38	0	2

AVIS GLOBAL

Le CLILLAC-ARP est une unité solide, disposant d'une stratégie scientifique claire et se caractérisant par son dynamisme, son attractivité et sa production scientifique de qualité dans des revues et chez des éditeurs réputés. Bien que ses thématiques de recherche soient assez éclatées (hétérogénéité qui se manifeste surtout dans son axe 1 et que met bien en évidence la rédaction du DAE), elles s'ordonnent toutes autour de la linguistique de corpus, de la linguistique expérimentale et, en général, du traitement automatisé ou semi-automatisé de données, avec un travail reconnu sur la langue anglaise. Le laboratoire déploie pour cela des compétences technologiques remarquables en matière de traitement du langage, une expertise dans la mise au point de protocoles expérimentaux, et une bonne sensibilité à l'interdisciplinarité.

Grâce à une politique de financement particulièrement énergique et à sa participation à de nombreux projets structurants (e.g. PAPTAN, Plateforme pour l'apprentissage profond pour la traduction automatique neuronale, projet SEPALE pour la prononciation de l'anglais), l'unité a réussi à se doter des ressources nécessaires pour mener à bien ses recherches, accueillir des chercheurs étrangers de renom, acquérir des équipements de pointe (infrastructures de stockage des données, serveur de calcul pour l'apprentissage profond, serveurs pour la traduction automatique, chambre sourde...) et offrir un environnement favorable à l'ensemble de son personnel, mettant un soin tout particulier à l'accueil des nouveaux collègues, des doctorants, fortement impliqués, et des jeunes chercheurs, français ou étrangers. Si l'attractivité de l'unité par la reconnaissance que lui confèrent ses succès en réponse à des appels aux projets compétitifs qui viennent d'être mentionnés est évidente au niveau local et national, elle reste cependant plus limitée au niveau international.

La production scientifique du CLILLAC-ARP est quantitativement appréciable et équilibrée, correctement répartie entre ses composantes ; elle est surtout de qualité, ce qui se traduit par sa participation à des revues à fort facteur d'impact (e.g. Journal of Phonetics, Journal of French Language Studies, International Journal of Learner Corpus Research) ou par l'accueil que des éditeurs prestigieux (comme John Benjamins, Oxford University Press, De Gruyter, Peter Lang) réservent à ses travaux.

L'unité a une productivité sérieuse en matière de corpus et d'outils informatiques, mais n'a pas de politique très explicite de développement, et plus généralement de gestion de ses interactions avec le monde non académique, alors qu'elle occupe un créneau très porteur et qu'elle peut s'appuyer sur une structure dédiée (Société d'Accélération du Transfert de Technologies — SATT — ORGANEO). Ses données sont rarement mises à la disposition du public.

Le CLILLAC-ARP s'est organisé en axes et en sous-axes, tout en définissant des axes transversaux, destinés à conjurer le risque d'atomisation de la recherche, mais cette structure reste peu lisible. Ce manque de lisibilité traduit peut-être une nécessité d'augmenter la cohérence et les points de rencontre entre les différentes thématiques. □ Bien qu'il connaisse des problèmes de gestion financière et d'encadrement technique, le laboratoire est organisé de manière efficace et transparente. Il est soucieux de l'intégrité scientifique de ses membres ainsi que de l'impact éthique et environnemental de ses travaux ; il souscrit aux objectifs de la science ouverte.

La restructuration programmée pour le prochain contrat devrait encourager le CLILLAC-ARP à infléchir son mode de fonctionnement, trop influencé par l'existence des UFR auxquelles il est lié, et à mettre sur pied des projets propres à renforcer son identité d'unité de recherche.

ÉVALUATION DÉTAILLÉE DE L'UNITÉ

A - PRISE EN COMPTE DES RECOMMANDATIONS DU PRÉCÉDENT RAPPORT

L'unité a globalement tenu compte des recommandations précédentes, sauf peut-être en matière de retombées dans le monde socio-économique et de structuration.

De ce côté, un effort a, certes, été fait puisque le nombre d'axes a été réduit à deux. Mais cette structure pose encore d'importants problèmes de lisibilité et d'efficacité.

L'axe 1 « Sémantique, Discours, Parole » est issu de la fusion des deux axes précédents « Discours, Parole, Cognition » et « Syntaxe, Sémantique, Pragmatique », correspondant aux deux grandes orientations des deux sous-axes « Parole et diversité » et « Syntaxe et Sémantique ». Les axes transversaux complétant les orientations de l'équipe sont problématiques en l'état : si « Comparaison, contraste, variation » est dédoublé dans chacun des deux sous-axes principaux, l'axe transversal « Discours et Société » fait figure de sous-axe, avec ses orientations propres. Ainsi, sa transversalité doit être clarifiée, tandis que l'autre axe transversal « Comparaison, contraste, variation » aurait pu bénéficier d'une présentation plus explicite en tant qu'élément intégré à l'axe principal, car l'intérêt commun pour l'analyse comparative pourrait jouer un rôle structurant. De son côté, l'axe 2 « Langues et Discours de Spécialité, Traductologie, Interculturalité » est désormais structuré en trois sous-axes, tandis que le quatrième sous-axe « Variation, normes, genres » s'incarne sous la forme d'un axe transversal autour des corpus. Ce dernier axe semble au final plus transversal et il pourrait être fusionné à l'axe transversal « Politique de la donnée et interopérabilité des corpus écrits et oraux », d'autant que cette transverse semble être tout à fait structurante pour l'unité, ayant favorisé la mise en place d'une synergie autour des corpus et de leur interopérabilité.

Enfin, comme le comité d'évaluation précédent l'a souligné, il aurait été bénéfique de mettre en place une structure thématique, en particulier pour l'axe 1, afin de favoriser la création de liens significatifs entre les travaux des membres et de promouvoir la collaboration autour d'objectifs communs. La présentation actuelle, trop énumérative, pourrait donner l'impression que ces recherches sont menées de manière isolée.

Le DAE rend compte de publications internationales, tant d'ouvrages et de chapitres de livres publiés chez des éditeurs internationaux de très bonne tenue que d'articles de revues internationales avec comité de lecture (on relève ainsi plusieurs articles publiés dans des revues particulièrement prestigieuses comme Journal of Phonetics, Journal of French Language Studies, International Journal of Learner Corpus Research). Ceci correspond bien à une recommandation émise en 2018. Mais ne disposant pas des données de production de la précédente évaluation, le comité n'est pas en mesure de proposer un comparatif chiffré.

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

L'unité a une stratégie scientifique claire, fondée sur l'utilisation de méthodes de la linguistique de corpus et de traitement de données en général, et sur une expertise en différentes langues, et en particulier en langue anglaise. Elle entend se saisir des développements méthodologiques et théoriques les plus récents et compte inclure méthodes expérimentales et approches interdisciplinaires dans ses orientations. Pour mettre en œuvre cette stratégie, l'unité utilise efficacement son environnement académique et partenarial principalement local.

Appréciation sur les ressources de l'unité

L'unité réussit à trouver principalement auprès de ses tutelles et de son environnement (Idex université Paris Cité, Labex Empirical Foundations of Linguistics) les ressources nécessaires pour mener à bien ses recherches, et offrir un environnement favorable à l'ensemble de son personnel permanent et non permanent, et cela bien qu'elle souligne que le développement de ses activités ces dernières années pose des problèmes tant de gestion financière que de soutien technique, en particulier pour renforcer ses travaux en linguistique expérimentale. Si de nombreuses données sont développées, elles pourraient être mieux mises en valeur.

Appréciation sur le fonctionnement de l'unité

L'unité et sa direction se sont organisées de manière efficace pour mettre en application les enjeux relevant des ressources humaines, comme la parité, la formation, la prévention. De même, elle prend en compte les enjeux relevant de la recherche durable, comme l'impact éthique (soumissions des projets au Comité d'éthique de la recherche de l'université Paris Cité), la prise en compte de l'impact environnemental, ou la préservation du patrimoine scientifique.

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

Points forts et possibilités liées au contexte

L'unité a défini et met en application une stratégie claire, largement partagée, reposant sur une expertise linguistique dans différentes langues et en anglais en particulier, ainsi que sur un intérêt marqué pour les corpus de données attestées et les méthodes de traitement de données. Dans cette perspective, elle développe des corpus et des méthodes empiriques dans l'ensemble de ses axes, en s'appuyant sur son axe transversal.

L'unité a pu pleinement exploiter son environnement immédiat pour mener au mieux sa stratégie. On peut citer l'apport de l'Idex (en particulier l'obtention d'un financement de la plateforme SYstran Pure Neural Server), du Labex EFL, mais aussi son partenariat avec le LLF et le LIMSI/LISN, ou avec la société SYSTRAN. Grâce à cela, l'unité a pu bénéficier de financements, mais a également pu mettre en place des synergies sur les thématiques de machine learning avec plusieurs partenaires. Elle s'est d'autre part positionnée sur des projets interdisciplinaires en se rapprochant de l'institut interdisciplinaire Data Intelligence Institute of Paris.

L'organisation de l'unité, avec des séminaires communs hebdomadaires, des conseils d'unité rassemblant l'ensemble du personnel où les orientations scientifiques sont décidées, ainsi que la présence d'un axe transversal portant la thématique de la gestion des données et de l'épistémologie de la science des données, permet de structurer l'ensemble des activités menées.

Points faibles et risques liés au contexte

La place primordiale des données et des corpus dans la stratégie de l'unité exige une plus grande prise en compte des impératifs de la science ouverte. En effet, dans la section « ressources » du site web, seuls deux corpus sont répertoriés, dont aucun n'est directement accessible (un lien est rompu et la mise sur Ortolang — Outils et Ressources pour un Traitement Optimisé de la LANGue — est différée dans l'autre cas). Or, dans le DAE, on relève des références à de très nombreux corpus construits ou en cours de construction au sein de l'unité.

On constate la place importante des UFR dans la vie du laboratoire (crédits, gestion, recrutements, doctorants, etc.) Ceci induit une structuration implicite qui n'est pas favorable à la définition d'une stratégie scientifique unifiée au niveau de l'unité, et qui pourrait être handicapante en termes d'identité et de visibilité.

L'unité pourrait étendre ses partenariats à d'autres acteurs académiques ou économiques. Ces nouveaux partenariats permettraient d'accroître les recherches sur contrat ANR (Agence Nationale de la Recherche) et Cifre (Convention industrielle de formation par la recherche ; une seule dans la période), d'obtenir ainsi davantage de contrats doctoraux, et d'attirer les meilleurs profils ; mais surtout ces nouveaux partenariats permettraient de mieux asseoir la maîtrise des nouveaux outils issus de l'Intelligence Artificielle comme les grands modèles de langue (de texte ou de parole), qui seront nécessaires à la linguistique de corpus du futur.

Dans les années à venir, l'unité souhaite développer ses travaux en linguistique expérimentale, mais cela nécessite un appui à la recherche plus important.

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

Malgré le rattachement de ses membres à trois UFR différentes, l'unité a su négocier des locaux pour tous ceux-ci. Elle a recherché dans son environnement (Idex université Paris Cité, Labex Empirical Foundations of Linguistics...) des moyens pour soutenir ses thématiques par des financements de stages, de doctorants et de post-doctorants. L'unité a pu compléter sa dotation récurrente de 72 k€ en moyenne sur la période par une moyenne de 130 k€ de ressources propres ; on note une augmentation significative de ce poste, la moyenne des ressources propres sur les trois dernières années étant de 223 k€. Elle a aussi pu dégager des financements internes pour favoriser l'émergence de nouvelles thématiques. Il est à noter qu'elle a réussi à obtenir un financement pour l'achat d'un serveur par l'Idex pour la plateforme Systran Pure Neural Server, permettant d'avoir accès à un système de traduction neuronale correspondant à l'état de l'art.

Points faibles et risques liés au contexte

Le nombre de projets partenariaux, qu'ils soient industriels ou publics (ANR — Agence Nationale de la Recherche) ou européens —, est limité (deux ANR sur la période, pour la même enseignante-chercheuse, et une seule Cifre — Convention industrielle de formation par la recherche —. Or les thématiques de l'unité peuvent permettre de mettre plus de projets en place. En particulier, il manque des projets d'ampleur — au niveau national ou européen —, qui seraient fédérateurs et structurants pour les différentes thématiques de l'unité.

En accord avec sa stratégie scientifique, l'unité a développé de nombreux corpus et bases de données ; cependant, le manque de visibilité d'ensemble de ces corpus ne permet pas d'évaluer précisément leur nombre et leur ampleur, et limite leur impact.

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

L'unité a une bonne parité homme/femme parmi les permanents, avec un léger déséquilibre (plus de femmes que d'hommes chez les doctorants). Les doctorants sont associés au conseil d'unité, et sont bien accompagnés, tant au point de vue environnement (ordinateur, bureau, financement de mission) qu'au point de vue scientifique (co-publications). L'unité est impliquée dans le suivi de carrière de ses personnels, qu'ils soient doctorants, maîtres de conférences ou PAR (Personnels d'appui à la recherche). Elle est consciente des risques encourus, que ceux-ci soient liés à la protection du patrimoine scientifique, au risque informatique ou aux situations d'urgence comme celle qu'a suscitée la pandémie de Covid-19 ; elle applique les mesures nécessaires, avec l'aide de ses tutelles.

Points faibles et risques liés au contexte

La stratégie de certaines thématiques implique l'emploi de modèles (en particulier les LLM, Large Language Models), qui sont très coûteux à développer du point de vue de leur impact carbone ; pouvoir mesurer l'impact de certaines techniques de deep learning, comme cela est suggéré dans le DAE, n'est qu'une étape. En effet, l'impact carbone, comme les autres critères de nature éthique doivent également être pris en compte lors de la stratégie scientifique : existe-t-il des méthodes plus frugales pouvant permettre des expériences équivalentes ? Existe-t-il des méthodes plus transparentes, plus explicables ?

DOMAINE 2 : ATTRACTIVITÉ

Appréciation sur l'attractivité de l'unité

L'unité a considérablement amélioré son attractivité durant le présent contrat, essentiellement grâce à une production scientifique de qualité. Elle a aussi renforcé sa politique d'accueil de jeunes chercheurs, grâce à une politique de financement très active et développé une culture de projets. Les ressources complémentaires obtenues ont permis de fédérer des recherches novatrices et d'accueillir des chercheurs étrangers de renom. Par ailleurs, l'unité s'est montrée davantage soucieuse du profilage « recherche » des emplois publiés, avec des résultats manifestes. Si l'attractivité de l'unité est manifeste au niveau local et national, elle reste cependant plus limitée au niveau international.

- 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

L'attractivité de l'unité est incontestable. Elle se manifeste d'abord par un rayonnement scientifique dû à une production scientifique de qualité. On peut citer ici les travaux sur la catégorie du nom, la détermination nominale et l'interface entre nom et discours. Les enseignants-chercheurs de l'unité ont ainsi mis en évidence une classe de noms jamais repérée auparavant : les noms polaires, dont la particularité est de pouvoir s'interpréter comme une interrogative polaire indirecte lorsqu'ils sont placés dans certains contextes syntaxiques permettant des interrogatives indirectes. Les travaux sur l'ellipse et l'anaphore, en français et en anglais, à partir de corpus et d'expériences psycholinguistiques d'acceptabilité, ont aussi renforcé incontestablement la bonne réputation du laboratoire. Il en va de même pour les études sur les schémas lexico-grammaticaux, en particulier en langues de spécialité, dans la mesure où elles interrogent frontalement l'articulation terminologie-phraséologie, ou pour celles menées dans le cadre d'une approche cognitiviste des grammaires de construction sur la double modalité dans des phrases du type *we might could solve this*. On peut aussi mentionner les travaux relevant de l'approche plus traditionnelle en phraséologie, notamment sur les expressions idiomatiques et les parémies, sans oublier les recherches sur les collocations génériques, pour lesquelles une journée d'études a été organisée en 2018 avec des chercheurs d'autres unités, notamment le Lidilem (Laboratoire de linguistique et didactique des langues étrangères et maternelles) de l'université Grenoble Alpes et l'université de Birmingham (Royaume-Uni).

En second lieu, malgré la crise sanitaire, l'unité s'est bien adaptée pour développer et mettre en œuvre de nombreuses actions pour continuer à assurer un rayonnement national et international, en étant particulièrement dynamique dans l'organisation de manifestations scientifiques : journées d'étude au niveau national dans le cadre de projets financés, ou colloques internationaux, notamment sur la traductologie, en collaboration avec l'université de Genève et l'université libre de Bruxelles, ainsi que la série de séminaires PCCM (Paris-Cologne Collaborative Meetings) organisés entre 2017 et 2021.

L'unité est aussi attractive par sa politique d'accueil et d'accompagnement d'enseignants-chercheurs venant d'autres laboratoires. Ainsi, chaque maître de conférences entrant dispose désormais d'une dotation de 10 k€ pour ses activités de recherche (achats de matériel, missions, etc.) et bénéficie en outre d'une décharge d'enseignement de 48 heures ETD. Cette politique permet aux nouveaux maîtres de conférences de développer dès leur arrivée des projets et collaborations à l'échelle internationale. Ils peuvent par ailleurs demander leur affiliation au Labex EFL, ce qui leur permet de développer leurs travaux au sein d'axes de recherche très structurés et de bénéficier ainsi d'une dynamique de recherche au sein d'un autre réseau de

chercheurs. L'accompagnement dans la carrière est également une des priorités de l'unité, qui encourage l'ensemble des enseignants-chercheurs à diffuser leurs travaux sur les plans national et international par une politique de financement des missions particulièrement active. Par ailleurs, la publication de profils « recherche » plus affirmés pour tel ou tel axe permet de renforcer significativement les effectifs de chaque équipe.

La politique d'accueil et d'accompagnement des doctorants et post-doctorants est elle aussi particulièrement attractive. Les doctorants sont encouragés à s'insérer rapidement dans un réseau international de recherche et bénéficient de conditions financières favorables pour acquérir du matériel de travail performant (ordinateurs, logiciels, etc.). Ils sont ainsi fortement encouragés à se former aux techniques de pointe et aux exigences de la recherche en suivant diverses formations (ateliers, écoles d'été, etc.). Sur le plan scientifique, doctorants et post-doctorants sont de plus en plus intégrés aux projets financés de l'unité, ce qui leur permet de se former par la recherche.

Le laboratoire est également investi dans l'accueil de chercheurs internationaux, par l'intermédiaire d'un programme international de « Chercheurs invités » pour de courtes périodes, afin qu'ils puissent intervenir dans les séminaires de l'unité. Il attire aussi des chercheurs lauréats de bourses de prestige (notamment une chercheuse de l'université de Californie à Davis, États-Unis, titulaire d'une bourse Fulbright) et il bénéficie encore de l'accueil de chercheurs sélectionnés par la Chaire internationale du Labex EFL.

L'attractivité de l'unité par la reconnaissance que lui confèrent ses succès en réponse à des appels à projets compétitifs est aussi évidente au niveau local et national (mais reste limitée au niveau international : voir infra). Le laboratoire a su profiter de divers financements par projets dans le cadre de plusieurs appels financés par l'Idex au niveau de l'université Paris Cité : projet SEPALE (Solutions pour l'enseignement de la phonétique appliquée aux langues étrangères), projet Son-Discours (From Sound to Discourse) entre autres ; on peut aussi citer ici le projet particulièrement important PAPTAN (Plateforme pour l'apprentissage profond pour la traduction automatique neuronale).

L'unité est encore attractive par la qualité de ses équipements et de ses compétences techniques. Elle a su développer l'aspect technique de la recherche en alliant acquisition d'équipements de pointe et développement des compétences des enseignants-chercheurs dans deux domaines : la linguistique de corpus (mise en place de serveurs, outils d'interrogation et plateformes pour le stockage) et la linguistique expérimentale, en psycholinguistique et phonétique-phonologie. Le laboratoire dispose ainsi d'une infrastructure de stockage des données (serveur CLILLAC-ARP), d'une plateforme d'interrogation de corpus sur laquelle sont installés de nombreux corpus en langues de spécialité, d'un outil d'annotation collaborative (installé sur serveur), d'un outil de textométrie en ligne, d'un serveur de calcul (PAPTAN) pour l'apprentissage profond, de deux autres serveurs SYSTRAN pour la traduction automatique, etc. Elle dispose enfin d'une chambre sourde, mutualisée avec le laboratoire LLF ; cette pièce contient des équipements de pointe (électroencéphalographe, oculomètre, échographe, électroglottographe) permettant de réaliser des expériences psycholinguistiques et de phonétique articulatoire.

L'unité a dans son ensemble une activité éditoriale soutenue (nombreuses éditions et co-éditions de volumes publiés par John Benjamins ou par Edinburgh University Press), et est impliquée dans l'expertise de projets de recherche et de dossiers de recrutements pour d'autres universités (en Suisse, en Belgique, au Québec, pour le Center for Translation Studies de l'université de Vienne) ainsi que dans l'évaluation de publications pour des revues et maisons d'éditions de renom (John Benjamins, De Gruyter, etc.).

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

L'attractivité de l'unité pour les appels à projets est essentiellement nationale.

Sa contribution à la construction de l'EER est difficile à évaluer. Si le volume de publications conjointes avec des chercheurs européens de renom est raisonnable (notamment avec les travaux mentionnés supra sur l'ellipse et l'anaphore parus dans les revues *Glossa* et *Corela*), le nombre de contrats formels avec la Commission européenne dans le cadre du programme Horizons 2020, ou avec d'autres institutions européennes de recherche et d'enseignement supérieur, reste limité pour un laboratoire qui compte près d'une quarantaine d'enseignants-chercheurs. On ne peut guère mentionner qu'une seule action Marie Skłodowska-Curie obtenue par un membre de l'unité. Le comité constate en particulier que les membres permanents et post-doctorants du CLILLAC-ARP ne candidatent pas auprès du Conseil européen de la recherche et de ses différents programmes, dont certains portent pourtant sur des thèmes couverts par leurs recherches, ni auprès d'autres agences de financement extra-européennes susceptibles d'être intéressées à soutenir des thématiques de l'unité.

Certes, plusieurs programmes européens Hubert Curien ont été portés par des membres du laboratoire au cours du contrat qui s'achève (e.g. avec la Slovénie en 2017 et 2019 pour la création de ressources terminologiques slovène-français d'aide à la traduction spécialisée ; avec l'Irlande en 2019 ; avec l'Estonie de 2019 à 2022 pour

étudier les questions en situation de surprise dans une perspective contrastive français/anglais/estonien ; avec l'Allemagne en 2020 et 2021 sur le même thème, dans une visée contrastive français/anglais/allemand). Mais ces programmes bilatéraux sont de faible envergure, et les financements qu'ils rapportent permettent seulement des courtes missions, et n'offrent donc pas la possibilité de soutenir solidement la recherche. En outre, le Comité note que les succès obtenus en réponse à ces appels d'offres sont le fait d'un groupe restreint de chercheurs.

DOMAINE 3 : PRODUCTION SCIENTIFIQUE

Appréciation sur la production scientifique de l'unité

Avec une production écrite et orale de 499 items, soit une moyenne de 2,6 par chercheur et par an, la production scientifique de l'unité est conséquente et proportionnée au potentiel ; elle paraît correctement répartie entre ses personnels. Sa qualité est confirmée par le renom des éditeurs scientifiques internationaux (e.g. Edinburgh University Press, John Benjamins, Oxford University Press, De Gruyter, Peter Lang...) et des revues (e.g. Journal of Phonetics, Journal of Laboratory Phonology, Review of Cognitive Linguistics, Journal of Linguistics...) accueillant ses productions, la liste extensive de colloques de référence, ainsi que le souci manifeste de l'intégrité scientifique, de l'éthique et de la science ouverte. Il faut y ajouter une productivité sérieuse en matière de corpus et d'outils informatiques.

- 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 écrite de l'unité s'élève à 252 items sur l'ensemble du contrat, et se répartit en 114 items pour l'axe 1 (15 permanents) et 138 pour l'axe 2 (20 permanents), soit une productivité similaire des deux axes, de l'ordre de 6,8 à 7,8 pour chaque chercheur, soit 1,3 par chercheur et par an. Les (co-)directions d'ouvrages et numéros de revue sont comparables, l'axe 1 se distinguant par une proportion plus forte d'articles tandis que l'axe 2 affiche autant d'articles que de chapitres d'ouvrages : une distribution conforme aux normes de ces deux grands domaines.

La qualité de ces productions ne fait pas de doute, comme en témoignent les lieux de publications. On relève en particulier pour l'axe 1 des ouvrages ou chapitres d'ouvrages chez Edinburgh University Press, John Benjamins, Oxford University Press, De Gruyter, Peter Lang. Quant aux articles dans les revues, on observe parmi celles-ci un certain nombre des plus réputées dans les domaines concernés : Journal of Phonetics, Journal of Laboratory Phonology, Review of Cognitive Linguistics, Journal of Linguistics, Journal of French Language Studies, International Journal of Learner Corpus Research (mais les bonnes revues nationales ne sont pas négligées : Anglophonia, Corela, Discours, par exemple). L'axe 2 n'est pas en reste avec ses publications chez John Benjamins, University of Chicago Press, Firenze University Press, Peter Lang, Bloomsbury Publishing, ou avec sa participation à des revues de réputation comparable (Revue Romane, Meta, European Journal of Language Policy, Fachsprache, International Journal of Learner Corpus Research, Revista de Filología Románica), comme à des revues plus nationales, ASp, Neologica.

La production écrite de l'unité est complétée par son pendant oral : 247 communications et posters (dont il ne serait pas simple d'établir la répartition par axe ; de même, la distribution de la production, tant écrite qu'orale, n'est pas spécifiée par sous-axes). Les productions orales sont en moyenne de 6,5 par chercheur (soit 1,3 par chercheur et par an). S'il paraît moins pertinent d'en citer certains plus que d'autres, les colloques fréquentés sont dans l'ensemble d'excellente tenue, avec une forte dimension internationale, sans que l'unité néglige pour autant son implication sur la scène nationale.

De façon générale cette production paraît correctement répartie entre les personnels, quand bien même certains présentent une productivité plus intense que la moyenne. On soulignera la politique volontariste de l'unité, aussi bien au titre de l'axe 1 que de l'axe 2, pour l'implication des doctorants, post-doctorants et jeunes maîtres de conférences dans cette activité, souvent en co-signature avec des chercheurs expérimentés, et ce dans les revues et colloques de plus grand prestige.

Compte tenu de la multiplicité de ses spécialités linguistiques, il serait difficile de se livrer à une analyse détaillée des positionnements théoriques et méthodologiques de l'unité. Toutefois, celle-ci s'est engagée dans une politique partagée qui doit être soulignée, l'amplification du recours à l'ancrage empirique avec recours aux technologies de pointe, et le développement déterminé de corpus et d'outils numériques afférents : participation au développement de TEICORPO, plateforme textométrique *l'rameur*, plateforme de collecte OCTAVES, projet SPECTRANS, NeuroViz, traduction automatique spécialisée PAPTAN, projet SEPALE pour la prononciation de l'anglais. Cette politique s'est ainsi traduite par l'implication dans une liste particulièrement consistante de projets et programmes ((PHC Ulysse, SEPALE, REMEDLANG, DLLA...)) et — ce qui est plus pertinent encore sans doute — s'inscrit dans une démarche d'épistémologie de la science des données linguistiques, et en particulier des jeux de données conçus comme objets privilégiés. Cette démarche fédérée par l'axe transversal à l'unité : « Politique de la donnée et interopérabilité des corpus écrits et oraux ».

Du point de vue des considérations éthiques et d'intégrité scientifique — y compris le RGPD (*Règlement Général sur la Protection des Données*) —, l'unité a recours au Comité d'Éthique et de Recherche de l'université Paris Cité, une instance externe qui en garantit l'objectivité. En interne, elle est particulièrement attentive aux co-signatures, notamment pour promouvoir l'implication de doctorants, et affiche une politique volontariste en ce qui concerne les préconisations de la science ouverte.

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

Avec une production écrite moyenne de 1,3 par chercheur et par an, il y a sans doute une marge de progression qui supposerait une politique plus incitative, quand bien même la qualité des lieux de publication et des publications elles-mêmes doivent continuer de primer.

Du point de vue du pilotage de l'unité, on remarque que les données fournies ne détaillent pas les axes et sous-axes dont elles relèvent : ce sont ainsi les données chiffrées de production écrite proposées dans le bilan de l'axe 2 qui ont permis de reconstituer celles de l'axe 1 ; pour la production orale, aucun élément ne permet — de façon commode en tout cas — d'aller au-delà des données brutes générales du document Excel. L'absence de ces éléments, du moins dans les documents fournis au comité, obère la possibilité d'un suivi précis, et des politiques incitatives que celui-ci pourrait permettre.

Enfin, mais ce n'est pas le laboratoire qui est ici en cause, si le recours au Comité d'éthique et de recherche de l'université Paris Cité est une excellente chose, l'unité elle-même souligne que le temps de réactivité dudit Comité est trop long pour un fonctionnement optimal.

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'unité est consciente des implications sociétales de ses thématiques. Plusieurs de ses travaux ont débouché sur des applications, principalement dans les domaines de la pédagogie des langues (e.g. phonologie des apprenants, jeux sérieux pour l'apprentissage de l'anglais, base d'aide à la rédaction de textes scientifiques...) et de la traduction (e.g. outils servant aux activités connexes à la traduction automatique, comme *Competence Awareness in TranslatiOn (CATO)*, corpus d'archives de traductions...). Les relations avec la criminalistique et la linguistique forensique sont également un point fort. Elle a aussi une intense activité de production de corpus et de plateformes. Mais les interactions avec le monde non académique, le développement de produits et les interventions dans les débats de société devraient être plus nombreux et variés. Elles ne semblent par ailleurs pas faire l'objet d'une stratégie, mais résultent plutôt d'effets d'aubaine.

- 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

De nombreuses thématiques de recherche du CLLAC-ARP — spécialement celles qui sont menées au sein de l'axe 2 — sont en soi prometteuses de retombées en direction des mondes social et économique ou ont une dimension sociétale évidente. Il en va ainsi des recherches menées en socio-phonétique, en acquisition du langage, en terminologie, en traductologie, en pédagogie, en analyses discursive et argumentative ou encore en interculturalité et en politique linguistique. L'unité est très consciente de ces implications. Bien mieux, de nombreux travaux très spécialisés menés au sein du laboratoire ont pour vocation de trouver des applications immédiates. On songe à l'élaboration de critères automatisés d'identification de la radicalisation sur internet, à la mise au point de programmes de deep learning, aux études ayant pour objet la phonologie des apprenants, à l'exploitation des jeux sérieux dans l'apprentissage de l'anglais, à la mesure de l'apport des langues et cultures autochtones aux enjeux socio-environnementaux contemporains ou à la recherche en interprétation en service public ou en médiation linguistique. Mais sur ces points, on en reste à la reconnaissance du potentiel, sans mise en œuvre concrète, ou alors de manière peu structurée.

Certains de ces travaux sont menés dans un cadre collectif : c'est le cas du projet SEPALE (Solutions pour l'enseignement de la phonétique appliquée à l'anglais), permettant l'apprentissage de l'alphabet phonétique international, du projet Pervasive game for health and languages integration, qui vise à l'intégration linguistique et culturelle des réfugiés et des demandeurs d'asile par le biais d'une sensibilisation à la santé ou encore du projet ARTES (Base d'Aide à la Rédaction de TExtes Scientifiques), ayant mené à la création d'une plateforme d'aide à la rédaction et à la traduction dans les domaines spécialisés, laquelle voit ses ressources constamment développées depuis 2010.

Tous ces travaux ont débouché sur des interactions avec le monde non académique et cela principalement dans deux directions : la pédagogie des langues et la traduction.

En ce qui concerne la pédagogie, on note la mise sur pied d'une journée d'étude sur la phonologie de l'anglais, destinée à un public constitué d'enseignants du secondaire et d'inspecteurs d'académie, la mise au point de formations médicales en anglais et en français, en vue de proposer un modèle d'apprentissage disciplinaire par problèmes, ainsi que plusieurs publications à visée pédagogique.

Dans le monde de la traduction, l'unité revendique des liens avec des associations professionnelles comme l'ENPSIT (European Network for Public Service Interpreting), l'AFFUMT (Association française des formations universitaires aux métiers de la traduction), la Société française des traducteurs, ISM interprétariat-Paris, ou avec des entreprises du secteur de la traduction automatique (RWS, SYSTRAN, memoQ, Wordfas) ainsi que des contacts avec l'association d'interprétation ISM (Inter Service Migrant) ou avec la Direction générale de la traduction de la Commission européenne

Mais d'autres collaborations plus ponctuelles sont à relever. Par exemple, l'expertise de plusieurs membres de l'unité en reconnaissance vocale a engendré des liens forts avec le Service National de Police Scientifique, et ces liens ont à leur tour débouché sur des activités en direction du grand public.

L'activité de l'unité se caractérise par une intense production de corpus et de plateformes, dont certaines sont apparemment exploitables en dehors d'elle. C'est le cas de PAPTAN (Plateforme pour l'apprentissage profond pour la traduction automatique neuronale), de SPECTRANS, ensemble d'outils servant aux diverses activités connexes à la traduction automatique, ou d'OCTAVES, corpus d'archives de traductions. L'équipe, particulièrement à travers un de ses membres, a aussi mis au point une série de logiciels de traitement automatique du langage, de traduction et de deep learning accessibles aux utilisateurs extérieurs intéressés.

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

L'unité estime consacrer 10 % de son énergie à l'expertise technique et à l'aide aux politiques publiques, dix autres pourcents à la dissémination des résultats de la recherche et 5 % à la valorisation de celle-ci. Ces estimations apparaissent surestimées, dans la mesure où l'unité compte au nombre des interactions avec le monde non académique d'une part le fait que ses recherches se caractérisent « par le recours à des données authentiques très actuelles » et d'autre part le fait qu'elles irriguent de nombreux enseignements de licence ou de master. Mais le premier point ne constitue pas un fait tangible d'interaction et de transfert, et le second relève davantage de la formation par la recherche que du développement.

Si l'unité est consciente des implications sociétales de ses thématiques de recherche, le paragraphe qu'elle consacre à cet ancrage social dans son DAE est hors sujet, car elle ne propose pas de stratégie clairement identifiable pour assurer ce transfert vers les mondes extra-académiques.

Il y a d'ailleurs une nette et surprenante disproportion entre les fortes potentialités des recherches menées dans l'unité et leurs retombées effectives : ainsi qu'on l'a relevé parmi les points positifs, un très grand nombre de travaux ont pour vocation de déboucher sur des produits ou sur des actions en direction de la société, mais on ne voit pas que la concrétisation de ces potentialités soit particulièrement recherchée. Il est ainsi étonnant que le DAE, sur ces aspects, ne mentionne jamais d'interaction avec les services supports dédiés de l'établissement s'il en existe, ni avec la SATT ERGANEO dont la tutelle est actionnaire et qui se doit d'assurer les activités de transfert, y compris en termes financiers, à partir des déclarations des membres de l'unité.

Un certain nombre de travaux ont certes des retombées pratiques, mais au sein du monde académique seulement. Par exemple, le projet « Solutions pour l'Enseignement de la phonétique appliquée aux langues étrangères » a effectivement débouché sur des solutions, mais qui ne profitent qu'aux enseignements d'anglais de l'université Paris Cité. De même, la mise en place de la plateforme SYSTRAN *Pure Neural Server* sur les serveurs de l'université permet d'optimiser un enseignement pragmatique de la traductologie, mais son accès est réservé aux membres de l'institution.

Le comité note surtout que les interactions de l'unité avec le monde non académique n'ont pas de caractère systématique. Non seulement il n'y a qu'un très petit nombre de membres qui se sont engagés dans cette direction, mais les actions menées ne semblent ni être collectivement planifiées ni être le résultat d'une réflexion stratégique d'ensemble. Ainsi, le DAE ne permet pas de comprendre dans quel cadre de valorisation et de transfert se font les interactions signalées. Par ailleurs, on peut être surpris du faible nombre de thèses Cifre dans ce secteur pourtant porteur.

L'activité de vulgarisation est également le fait de quelques-uns de ses membres seulement (comme la directrice de l'unité, ou un doctorant qui ne totalise pas moins de 14 interventions dans les médias). Et l'absence de caractère systématique se note ici aussi : l'unité semble se contenter de répondre à des demandes ponctuelles sur les questions qui font débat dans l'espace public et ne pas avoir sur ce point de politique concertée. La visibilité de l'unité en tant que telle en souffre nettement.

ANALYSE DE LA TRAJECTOIRE DE L'UNITÉ

Comme indiqué dans l'historique de l'unité, celle-ci est née de deux fusions, ayant abouti à une organisation en axes et sous-axes assez touffue.

La dénomination de CLILLAC-ARP ne rendant plus exactement compte des thématiques de recherche actuelles de l'unité (par exemple, la lexicologie n'y a plus guère de place comme telle), cette dernière a résolu de se donner un nouveau nom pour le prochain contrat : Approches Linguistiques Théoriques, Appliquées et Expérimentales : langues et cultures connectées (ALTAE).

Poursuivant son entreprise de structuration, et pour tenir compte de l'émergence de nouvelles thématiques au cours du contrat écoulé, l'unité se propose d'organiser le travail à venir autour de quatre modules cadre (dont trois sont appelés « équipes » et l'un « atelier permanent ») donnant lieu à une subdivision plus fine en treize items appelés « actions thématiques » et « opérations ». Les trois équipes sont les suivantes :

- Enjeux actuels de la Traduction et Communication des Cultures en Contact;
- Dynamique des Discours et Défis Sociétaux;
- Phonétique, Acquisition et Psycholinguistique.

L'unité identifie ainsi clairement trois champs disciplinaires distincts et cohérents, qui s'émancipent désormais en partie de la frontière historique entre UFR au profit d'une logique scientifique.

Suivant en cela les incitations à identifier et intégrer l'émergence de nouvelles thématiques, la partie projet du document ne détaille pas moins de dix actions thématiques (par équipe, respectivement 3, 4 et 3). Le volume même de la rédaction consacrée à ces « actions thématiques », comparé au caractère résumé de la présentation des équipes elles-mêmes, incite à considérer que ces sous-équipes, relativement indépendantes en dépit de quelques participations croisées, constituent de fait la véritable organisation scientifique de l'unité. Un tel degré de dispersion, pour un effectif total de 38 enseignants-chercheurs, interroge : il mérite sans doute d'être davantage précisé, voire reconsidéré.

Cette situation n'est pas surprenante au vu de l'historique de l'unité, et l'on perçoit que celle-ci s'est employée à y faire face en incitant l'ensemble de ses membres à se focaliser sur la question des corpus, et cela depuis au moins 2013, comme en témoignent les structures successives suivantes :

2014-2018 Axe transversal : Épistémologie et Interopérabilité des Corpus ;

2019-2023 Axe transversal : Politique de la donnée et interopérabilité des corpus écrits et oraux ;

2024- Atelier Permanent de Réflexion Épistémologique et Méthodologique.

Le positionnement de cet Atelier au centre de la représentation graphique de l'unité, la volonté de l'inscrire dans le cadre d'une réflexion épistémologique spécifique des sciences du langage, ou plus prosaïquement le fait que la description détaillée du projet commence par cet Atelier, et fait, elle, l'objet d'une présentation générale à la fois ample et précise, constituent autant de manifestations de l'effort continu de l'unité pour construire une identité scientifique distinctive.

En l'état, il n'est toutefois pas certain que cela suffise à définir l'originalité de l'unité dans le champ des sciences du langage en France, et il est sans aucun doute souhaitable qu'elle poursuive son entreprise de cohésion, notamment par l'identification et le développement d'interactions lisibles entre équipes sur des thématiques proprement linguistiques. Une mise en valeur de quelques projets phares pourrait ainsi constituer un facteur d'identité collective.

Les interactions avec les autres unités de recherche sur le même site, ou plus généralement au niveau local, doivent aussi être davantage pensées et structurées.

RECOMMANDATIONS À L'UNITÉ

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

Il conviendrait que l'unité consente les efforts nécessaires pour mettre en adéquation les impératifs de science ouverte, qu'elle défend à juste titre, et la diffusion et la valorisation des nombreux corpus et jeux de données qu'elle a développés ou qu'elle continue de développer ; cela peut passer par la mise en œuvre de plans de gestion des données, même quand le financeur ne l'exige pas, en prenant les mesures nécessaires pour que ces corpus soient visibles par la communauté scientifique sur les plateformes telle que Ortolang.

L'unité est encouragée à réfléchir collectivement à mettre en place des projets collaboratifs d'ampleur, nationaux ou européens, qui seraient fédérateurs et structurants pour les différentes thématiques de l'unité.

Sur les thématiques où la linguistique outillée nécessite l'usage des nouveaux outils liés à l'IA, il serait profitable d'étendre les partenariats, tant académiques que privés. Cela permettrait une meilleure réactivité, l'évolution de ces outils étant extrêmement rapide, et conduirait à de nouveaux projets, augmentant les ressources propres de l'unité et les possibilités de financement de doctorats et post-doctorats. Cela favoriserait également le recrutement d'un enseignant-chercheur avec un profil spécifiquement en TAL, fondamental pour asseoir cette thématique au sein du laboratoire de manière pérenne.

Enfin, l'unité devrait compléter ses efforts sur la recherche durable avec une réflexion éthique sur l'utilisation de ces outils dans la société, dans les thématiques relevant de sa stratégie.

Recommandations concernant le domaine 2 : Attractivité

Le comité recommande de développer et d'élargir les réponses aux appels à propositions aux niveaux européen et international, ce qui renforcerait encore son attractivité. Ces projets incluent en effet des préciputs importants pour les structures d'accueil, ce qui permet à celles-ci de mener des opérations de recherche dans des thématiques scientifiques qui ne relèvent pas toujours des projets retenus. Cette recommandation vaut pour les enseignants-chercheurs de l'unité, mais aussi et surtout pour les doctorants et post-doctorants, qui doivent être fermement encouragés à chercher des financements complémentaires pour soutenir leur recherche.

En particulier, les compétences de certains membres permanents et post-doctorants du CLILLAC-ARP devraient les conduire à candidater auprès du Conseil européen de la recherche (ERC – European Research Council) et de ses différents programmes (Starting, Consolidator, Advanced). D'autres agences de financement internationales et extra-européennes pourraient aussi être intéressées par certaines thématiques de l'unité : aux États-Unis (NSF, National Science Foundation), au Japon (JSPS, Japan Society for the Promotion for Science), à Hong Kong (RGC, Research Grant Committee), ou à Taiwan (MOST, Ministry of Science and Technology)

Enfin, une réflexion est attendue sur la politique de l'unité quant aux moyens disponibles au sein du Grand Équipement Documentaire (GED) ouvert sur le Campus Condorcet.

Recommandations concernant le domaine 3 : Production scientifique

Le Comité recommande à l'unité de conduire une politique incitative pour le développement des productions écrites et orales, dont le volume peut encore être augmenté sans exercer une pression excessive sur ses membres.

Pour un pilotage fin de l'unité, il conviendrait de mettre en place un suivi des productions détaillé par équipes et par action thématique qui puisse être facilement croisé avec les productions individuelles, et d'inclure dans ce suivi un outil permettant d'identifier séparément l'activité des doctorants de celle des enseignants-chercheurs.

L'unité pourrait demander à sa tutelle de renforcer les moyens du Comité d'éthique et de recherche, afin de permettre une réactivité plus en phase avec les besoins des équipes de recherche.

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

Les interactions avec le monde non académique et le développement de produits devraient être plus nombreux et couvrir un plus large éventail. L'exploitation des plateformes mises au point pourrait être une des pistes à explorer, et certains produits destinés au monde académique pourraient être exploités en dehors de ce dernier.

Le comité recommande que lesdites interactions avec le monde non académique fassent l'objet d'une politique explicite et systématique, dont les principes et les méthodes devraient être débattus collégalement à l'échelle de toute l'unité, politique qui assurerait une mise en œuvre concertée. Un des principes à retenir pourrait être que chaque recherche ayant potentiellement des retombées soit analysée dans la perspective de ses possibles applications dans le monde socio-économique. Pour ce faire, il conviendrait de s'entourer des compétences en matière de valorisation, de transfert et de R&D partenariale présentes dans les services supports de l'établissement et de la SATT ERGANEO, structure de valorisation dont la tutelle est actionnaire.

L'activité de vulgarisation et d'intervention dans les débats de société pourrait être plus dynamique et plus collégalement assumée, mais surtout plus systématique elle aussi : se reposer sur l'amélioration en cours d'un service de communication au niveau facultaire ou sur un site web renouvelé n'est pas une mesure suffisante.

DÉROULEMENT DES ENTRETIENS

DATES

Début : 10 décembre 2023 à 20h30

Fin : 11 décembre 2023 à 16h45

Entretiens réalisés en présentiel

PROGRAMME DES ENTRETIENS

Le 10 décembre, 20h30 : Rencontre préparatoire de la visite en présence du conseiller scientifique.

Le 11 décembre, 9 h 30 : Accueil (bâtiment Olympe de Gouges, 8 rue Albert Einstein 75013 Paris, salle 715

10h-10h30	Réunion à huis-clos avec les tutelles (Salle 715) : Mathieu Arnoux et Aouatif El Fakir .
10h30h-11h20	Session plénière (salle 720). Présentation du laboratoire : Natalie Kübler directrice de l'unité ; Avenir du laboratoire : Mojca Pecman, future directrice.
11h20-11h35	Pause.
11h35-12h00	Réunion avec l'axe 1 : Sémantique, Discours, Parole (salle720) : Agnès Celle, co-responsable de l'axe et directrice adjointe.
12h00-12h25	Réunion avec l'axe 2 : Langues et Discours de Spécialité, Traductologie, Interculturalité (salle720) : Natalie Kübler, responsable de l'axe et directrice.
12h25-13h30	Déjeuner (salle 715).
13h30-13h55	Réunion avec l'axe transversal : Politique de la donnée et interopérabilité des corpus écrits et oraux (salle720)) : Nicolas Ballier responsable de l'axe.
14h00-14h30h	Réunion à huis-clos avec les doctorants (salle720). Introduction par Marie Bouchet et Coline Caillol, représentantes des doctorants au conseil de laboratoire.
14h30-15h00	Réunion à huis-clos avec le personnel d'appui (salle 720) : Corinne Hamel responsable administrative, Loïc Liégeois, IGR, Brice Bricaud, IGE développeur.
15h00 -15h15	Pause.
15h15-15h45	Réunion à huis-clos avec la direction de l'unité, actuelle et future (salle 720) : Natalie Kübler, directrice, Agnès Celle, directrice adjointe, Mojca Pecman, future directrice, Emmanuel Ferragne, futur directeur adjoint.
15H45-16h45	Réunion à huis clos du comité d'experts en présence du conseiller scientifique (salle 715).
16 h. 45	Fin de la visite.

OBSERVATIONS GÉNÉRALES DES TUTELLES

Le Président

Paris, le 15 février 2024

HCERES
2 rue Albert Einstein
75013 Paris

Objet : Rapport d'évaluation de l'unité DER-PUR250024293 - CLILLAC-ARP - Centre de Linguistique Interlangue, de Lexicologie, de Linguistique Anglaise et de Corpus, Atelier de Recherche sur la Parole

Madame, Monsieur,

L'université Paris Cité (UPCité) a pris connaissance du rapport d'évaluation de l'Unité de Recherche **CLILLAC-ARP - Centre de Linguistique Interlangue, de Lexicologie, de Linguistique Anglaise et de Corpus, Atelier de Recherche sur la Parole**.

Présidence

Référence

Pr/DGDRIVE/2023

Affaire suivie par

Christine Debydeal -
DGDRIVE

Adresse

85 boulevard St-Germain
75006 - Paris

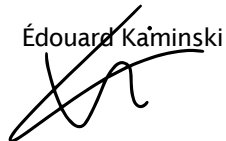
Ce rapport a été lu avec attention par la direction de l'unité (de la part de laquelle vous trouverez ci-joint un courrier listant les erreurs factuelles à corriger), la vice-doyenne Recherche et le doyen de la Faculté Sociétés & Humanités d'UPCité (de la part desquels vous trouverez également ci-joint un courrier à votre intention), par la vice-présidente Recherche d'UPCité et par moi-même.

J'adresse mes remerciements au comité HCERES pour la qualité du rapport d'évaluation, et vous indique ne pas avoir d'observations de portée générale complémentaires à apporter.

www.u-paris.fr

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

Édouard Kaminski



Les rapports d'évaluation du Hcéres
sont consultables en ligne : 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

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

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