

## SECTION 34

---

### SCIENCES DU LANGAGE

#### Composition de la section

Serge PINTO (président de section) ; Isabelle ROY (secrétaire scientifique) ; Pius Wuchu AKUMBU ; Angélique AMELOT ; Heather BURNETT ; Lieven DANCKAERT ; Rodolphe DEFIOILLE ; Sascha DIWERSY ; Maia DUGUINE ; Margherita FARINA ; Richard FAURE ; Nicola LAMPITELLI (membre du bureau) ; Alda MARI ; Philippe MULLER ; Léa NASH (membre du bureau) ; Tatiana NIKITINA (membre du bureau) ; Sophie POYDENOT D'ORO DE PONTONX ; Rudolph SOCK ; Eva SOROLI ; Alice VITTRANT. Anciens membres contributeurs : Rachid RIDOUANE ; William SAYER ; Ur SHLONSKY.

#### Résumé

Notre rapport décrit tout d'abord les unités mixtes de recherche de la section, leurs personnels, avec une attention particulière à la démographie des personnels statutaires. Il met en lumière la dynamique de recrutement au concours CNRS entre les rapports précédent et actuel. Il souligne l'importance de l'interdisciplinarité dans les sciences du langage et identifie les interactions phares entre sous-disciplines. Ces interactions sont présentées à travers des regroupements thématiques tels que la linguistique fondamentale, les approches transversales, et les variations et diversités. Le rapport de conjoncture, étape cruciale du mandat des sections du Comité national de la recherche scientifique (CoNRS), vise aussi à analyser les enjeux actuels et les perspectives de recherche. La dernière section de ce

rapport, construite sur la base d'un dialogue avec les directeurs et directrices d'unité, identifie les obstacles, les défis et les ouvertures auxquels la communauté scientifique des sciences du langage est confrontée.

---

#### Introduction

---

##### A. Les sciences du langage

Le « langage » est un objet de recherche commun à une multitude de sciences qui l'appréhendent en tant que capacité humaine universelle, intimement liée à des processus cognitifs et à des processus sociaux, mais

aussi en tant que système de signes (une langue) permettant l'expression de la pensée et la communication. Ces diverses sciences représentent autant de disciplines, sous-disciplines, domaines ou sous-domaines de recherche, qui au fil du temps ont dessiné une mosaïque d'approches méthodologiques et théoriques toutes complémentaires les unes aux autres : les sciences du langage sont plurielles, et sont centrales dans l'éventail des sciences humaines et sociales.

Rendre compte des contributions de ces différentes sous-disciplines au sein des sciences du langage peut prendre différentes formes en fonction de leurs considérations, en témoignent les divers découpages ou regroupements thématiques des précédents rapports de conjoncture du Comité national de la recherche scientifique (CoNRS). Sans innover, en nous fondant sur les rapports précédents, tout en nous prêtant à l'exercice de la synthèse et la concision, nous avons adopté un découpage en trois grands axes, qui seront développés de manière plus précise dans la partie IV « État des lieux scientifique des sciences du langage » :

- l'étude des caractères structurels du langage ainsi que des processus généraux qui en régissent l'emploi, dans la communication humaine, font l'objet notamment des recherches en linguistique fondamentale (phonétique, phonologie ; morphologie ; syntaxe ; sémantique ; pragmatique, philosophie du langage ; textes et discours ; lexicologie, lexicographie, traductologie) ;
- les approches transversales (linguistique de corpus, TAL, linguistique informatique ; sociolinguistique, linguistique anthropologique ; psycholinguistique, neurolinguistique, acquisition ; histoire des théories linguistiques, épistémologie), qui peuvent s'associer aux domaines de la linguistique fondamentale, utilisent des méthodologies différentes, souvent outillées et/ou expérimentales, dont certaines peuvent déboucher sur des applications concrètes dans l'industrie, la santé, ou l'éducation ;
- l'universalité du langage va de pair avec la diversité des langues naturelles. L'une

comme l'autre sont des sujets de recherche importants des sciences du langage, explorant ainsi les notions de variation, de diversité (évolution du langage ; linguistique historique ; typologie, diversité, variation linguistique ; dialectologie, contacts de langues ; langues signées) et de multimodalité.

## B. Une (brève) histoire de la section 34 et des sciences du langage au CNRS depuis 1949

Les « sciences du langage » sont l'expression d'une discipline relativement jeune, qui mûrit avec le temps et affine son périmètre de recherche autour de son objet d'étude depuis un peu plus de quarante ans. Dans un récapitulatif réalisé par le secrétariat général du Comité national (juillet 2010), l'évolution des intitulés et mots-clés des sections et commissions interdisciplinaires nous permet de suivre cette évolution pour la section 34 (cf. figure 1) et d'en dégager certains événements phares :

- **l'apparition des « sciences du langage » (section 42) date du mandat qui a débuté en 1983** : auparavant, une section centrée sur la linguistique générale, les langues et littératures étrangères couvrait le champ de recherche, aux côtés de sections distinctes tournées vers les études linguistiques, littéraires françaises, et les langues et civilisations (classiques et orientales) ;
- **les approches transversales ont intégré/réintégré le périmètre des sciences du langage au fil du temps** : si l'ethnolinguistique, la sociolinguistique, la psycholinguistique et la relation linguistique/informatique ont été intégrées dans les mots-clés de la section entre 1983 et 1991, on les retrouve dans d'autres sections ; par exemple, psycholinguistique et neurolinguistique en section 29 (fonctions mentales, neurosciences intégratives, comportements) et ethnolinguistique en section 38 (unité de l'homme et diversité des cultures) entre 2000 et 2004 ; entre 2004 et 2008, la section 27 (comportement,

cognition, cerveau) est créée, avec disparition de la psycho- et de la neurolinguistique et la mention du « langage » dans ses mots-clés, alors que la philosophie du langage apparaît en section 35 (philosophie, histoire de la pensée, sciences des textes, théorie et histoire des littératures et des arts) ; l'ethnolinguistique, quant à elle, demeure en section 38 (savoirs et cultures, approches comparatives) ; la section 07 (sciences et technologies de l'information : informatique, automatique, traitement du signal) apparaît en 1992, intégrant notamment dans ses mots-clés « intelligence artificielle » et « communication homme-machine », pour ensuite ajouter progressivement les termes « parole, langue naturelle » ;

- en 1992, la section acquiert le numéro 34 et s'intitule « Représentations, langages, communication » entre 1992 et 2000, puis « Langues, langage, discours » entre 2004 et 2012, pour enfin retrouver le titre de « Sciences du langage » à partir de 2012 ;
- pour répondre à une évolution des pratiques et au développement de l'interdisciplinarité, la **commission interdisciplinaire (CID) « Cognition, langage, traitement de l'information, systèmes naturels et artificiels » est créée en 2003 et subsiste jusqu'en 2012** (CID 45 de 2003-2008, sections concernées : 07, 29 et 34 ; CID 44 de 2008-2012, sections concernées : 01, 07, 27, 34, 36 et 40), ce qui a permis d'augmenter le recrutement de chercheurs et chercheuses en sciences du langage ;
- dès 2012, les mots-clés de la section 34 « Sciences du langage » se stabilisent, avec tout de même la circulation de quelques-uns d'entre eux entre la 34 et d'autres sections : par exemple, entre 2012 et 2016, les mots-clés « traitement automatique des langues et de la parole » apparaissent sous la section 07, « psycholinguistique et philosophie du langage » sont associés à la section 26 ;
- le périmètre de la section 34 prend sa **forme actuelle** en 2016 avec les mots-clés suivants :

- linguistique fondamentale : phonétique, phonologie, morphologie, syntaxe, sémantique, lexique ;
- signification et usage du langage : philosophie du langage, pragmatique ;
- typologie : universaux et diversité ;
- changement linguistique et évolution des langues : linguistique historique, contact de langues, linguistique anthropologique ;
- interaction et langage : sociolinguistique, analyse discursive, textes ;
- linguistique computationnelle : traitement automatique des langues, traitement du langage naturel, modélisation et simulation de phénomènes linguistiques ;
- psycho- et neurolinguistique : approches expérimentales et cliniques du langage ;
- histoire des théories linguistiques.

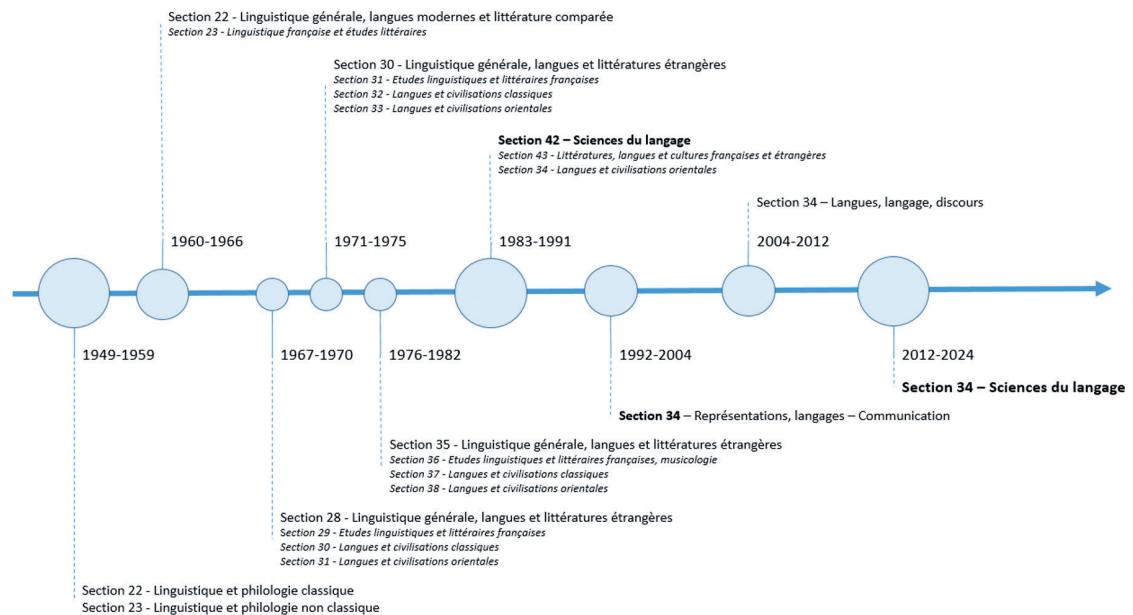
Si les sciences du langage partagent une partie de leur périmètre avec les sciences cognitives, les neurosciences, l'informatique et les sciences de l'éducation, elles gardent une spécificité liée à la place première qu'elles accordent au langage et aux langues. La proximité avec ces autres disciplines, du fait notamment de la constante augmentation de pratiques pluridisciplinaires et interdisciplinaires, entraîne une évolution notable vers des approches expérimentales, la généralisation de l'utilisation de gros corpus, de modèles informatiques et de bases de données.

En 2025, la numérotation des sections du CoNRS évolue et la section sciences du langage devient la section 36.

## C. Les enjeux actuels des sciences du langage

Comme toutes les sciences où l'approche expérimentale occupe une place grandissante, les sciences du langage ont été confrontées à une crise de la reproductibilité des résultats notamment en psycholinguistique, en

## Comité national de la recherche scientifique



**Figure 1 :** Évolution de l'intitulé de la section 34 et de ses prédecesseurs de 1949 à 2024.

acquisition du langage, et en neurolinguistique, mais aussi à un problème de qualité des ensembles de données dans beaucoup de tâches en traitement automatique des langues. La restructuration dans le milieu de l'édition scientifique vient se rajouter à celles-ci : la multiplication des revues en accès libre, pour lesquelles le paiement de frais de publication souvent élevés et le caractère opaque des arbitrages d'articles sont des exemples nous obligeant à repenser nos pratiques de diffusion de la recherche et adapter notre activité face au déploiement de la science ouverte.

La recherche en sciences du langage s'attache à préserver les capacités réflexives et critiques sur toute question impliquant la fonction, les usages et les effets des langues dans la société afin de contribuer à une information documentée et à des prises de décision éclairées. L'avènement de l'IA, où le traitement des données langagières occupe une place centrale, est probablement l'un des points ayant remis en question et changé en profondeur la perception et la relation du grand public à la génération automatique de textes, contenus et autres productions linguistiques.

Les systèmes fondés sur l'apprentissage automatique sont en plein essor, avec une qualité grandissante dans le traitement de ces données, au moins pour certaines langues. L'IA contribue d'ores et déjà au renouvellement profond des questionnements en sciences du langage (par exemple, quelle est la place de la diversité linguistique ? Quid des langues qui ne font pas partie des modèles d'entraînement des systèmes génératifs ? Dans quelle mesure l'IA façonne, homogénéise ou efface notre productivité langagière ? etc.) ; c'est aussi dans l'immédiat une opportunité pour ces dernières de disposer de nouveaux instruments qui permettent d'aborder ces nouvelles questions.

Un autre défi est représenté par les approches longitudinales dites *lifespan*. Celles-ci sont de plus en plus nombreuses et s'intègrent maintenant de manière naturelle dans les études en sciences du langage. À l'instar des approches synchronique/diachronique d'une langue, les études du langage chez le nourrisson (acquisition) et la population plus âgée (perte, trouble, remédiation) appréhendent l'évolution de la capacité langagière de manière ontogénique.

Enfin, la nécessité d'une attention particulière vers les langues peu dotées (régionales, des signes ou en danger) s'impose, en intégrant par exemple une approche expérimentale (TAL et autres), et en prenant en compte les bouleversements générés par les déplacements de populations et les contacts des langues qui en découlent.

De manière générale, les enjeux actuels et futurs en sciences du langage relèvent de l'approfondissement de l'interdisciplinarité tout en préservant et renforçant les fondamentaux disciplinaires issus de la linguistique fondamentale (cf. partie V). Le défi est de conduire les sciences du langage vers une pratique « avec et pour la société », déployant une valorisation environnementale et sociétale autant que faire se peut.

## D. Structure de notre rapport de conjoncture

Après avoir présenté les unités mixtes de recherche rattachées de manière principale à notre section (distribution géographique et thématique), nous avons souhaité décrire les personnels qui les composent, y compris les personnels non CNRS. Nous avons tenté d'identifier le plus précisément possible la démographie des personnels statutaires de chaque unité (chercheurs et chercheuses CNRS, enseignantes-chercheuses et enseignants-chercheurs universitaires, ITA, etc.). Il nous est apparu important également, afin de rendre compte de la trajectoire thématique de notre section, de rappeler la dynamique de recrutement au concours CNRS entre ce rapport et le précédent (2019-2023).

L'interdisciplinarité étant importante dans l'activité des sciences du langage, nous avons tenté d'identifier les interactions phares entre sous-disciplines, en guise d'introduction à notre lecture de la conjoncture des disciplines des sciences du langage au travers de regroupements thématiques déjà évoqués plus haut :

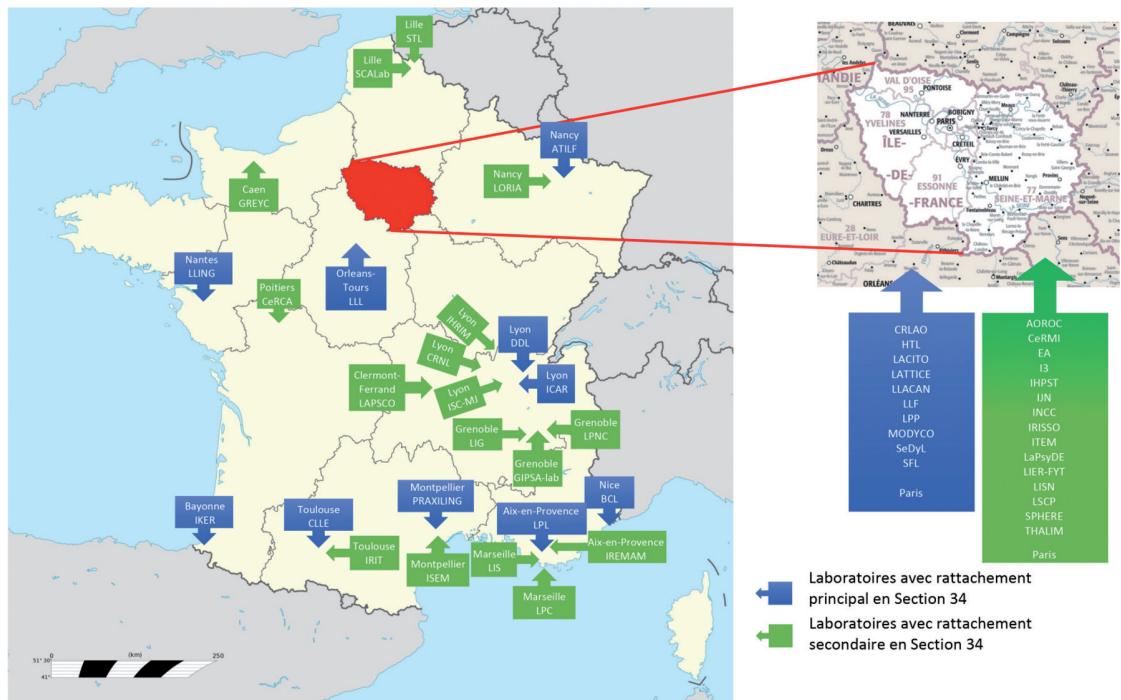
linguistique fondamentale, approches transversales, variations et diversités.

La rédaction du rapport de conjoncture est une étape importante lors du mandat des sections du Comité national de la recherche scientifique. Ce rapport a pour objectif « d'analyser les **enjeux** actuels de la recherche et les **perspectives** ouvertes par les principales découvertes de leur discipline ou groupe de disciplines ». Ainsi, nous terminons ce rapport par une dernière partie intitulée « Verrous actuels, défis prospectifs et thématiques émergentes ». La section 34 s'est inquiétée de répondre seule à l'ancrage – dans le temps et dans l'espace – de ses disciplines. Ainsi, nous avons contacté les directeurs et directrices d'unités rattachées de manière principale ou secondaire à notre section, pour les inviter à engager avec nous un dialogue, afin d'identifier de manière commune les verrous et les obstacles, mais également les défis et les ouvertures auxquels notre communauté scientifique fait face. Forts de ces échanges, nous avons surtout souhaité être au plus proche de la réalité des laboratoires dans les constats et propositions dont nous nous proposons de rendre compte dans notre rapport de conjoncture. Et ainsi établir des éléments prospectifs – tournés vers l'avenir – que nous aurions déterminés ensemble.

## I. Les unités mixtes de recherche et leurs personnels

Notre synthèse s'appuie sur les données issues de plusieurs sources : l'annuaire du CNRS (<https://www.cnrs.fr/fr/annuaires-du-cnrs>) et les informations du personnel disponibles *via* l'application Réséda. Sur ce dernier point, nous souhaitons remercier les directeurs et directrices d'unité, ainsi que les gestionnaires de laboratoires, qui nous ont permis de pouvoir compiler ces informations. La distribution géographique des laboratoires rattachés de manière principale et secondaire, telle qu'en septembre 2023, est présentée sur la

## Comité national de la recherche scientifique



**Figure 2 :** Distribution géographique des laboratoires rattachés à la section 34 (septembre 2023).

figure 2. Certains éléments sont déjà, ou seront prochainement, obsolètes du fait de changements de rattachements d'unités, d'affectations de chercheurs, ou autre, survenus au 1<sup>er</sup> janvier 2024 (cf. annexes 2 et 3). Cette photographie des ressources humaines demeure tout de même informative sur la période considérée.

## A. Les unités mixtes de recherche rattachées à la section 34

Les unités mixtes de recherche (UMR) affiliées de manière principale à la section 34 sont au nombre de 20 et sont présentées dans l'annexe 2. De manière corollaire, mais sans entrer dans les détails toutefois, la liste des 16 UMR rattachées de manière secondaire à la section 34 est donnée dans l'annexe 3; le

nombre de chercheurs et chercheuses CNRS de la section 34 affectés dans ces unités est indiqué.

Le périmètre scientifique de ces unités, sur la base de leur positionnement dans les mots-clés de la section à l'issue d'une consultation auprès de ces mêmes unités en septembre 2023, est visible sur la figure 3. De manière attendue, les laboratoires aux effectifs les plus importants couvrent un large éventail de disciplines thématiques. Ce sont souvent des laboratoires hors Île-de-France qui jouent pleinement leur rôle fédérateur d'animation de la recherche. Les unités centrées sur un nombre plus restreint de thématiques de recherche sont de deux types : unité rattachée de manière principale à la section 34 avec des effectifs moindres; ou bien unité rattachée de manière secondaire à la section 34, avec quelques chercheurs regroupés au sein d'une équipe de recherche tournée vers le langage.

RATT	Labo	N° DISC	P	H	P	M	O	R	L	S	E	M	P	D	I	S	T	H	P	S	A	C	O	H
			PHONET	PHONO	MORPHO	LEX	SEMANT	PRAG	DISCOURS	PHILO	EVO	TYPO	PHILO	DISCOURS	PHILO	TYPO	HIST-COMP	SOCIO	PSY-NEURO	ACQUIS	TAL	CORPUS	HIST-LING	
1	CLLE	18	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	DDL	16	Oui	Oui	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	STL	16	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	ATILF	15	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	LLF	15	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	LACITO	14	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	LLACAN	14	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	LLING	14	Oui	Oui	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	CRLAO	13	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	LLL	12	Oui	Oui	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	LPL	12	Oui	Oui	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	GIPSA-lab	12	Oui	Oui	Oui	Oui	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	
1	SEDYL	11	Non	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	Lattice	11	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	Praxiling	11	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	ICAR	9	Non	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	LPC	8	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	GREYC	8	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	LISN	7	Oui	Oui	Non	Non	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	Cemri	6	Non	Non	Oui	Oui	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	SFL	6	Non	Oui	Oui	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	EA	6	Oui	Non	Non	Non	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	IHRIM	6	Non	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
1	HTL	5	Non	Non	Oui	Oui	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	CeRCA	4	Non	Non	Non	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	
2	ISC	4	Non	Non	Non	Non	Oui	Non	Non	Oui	Non	Oui	Non	Non	Non	Non	Non	Oui	Oui	Non	Non	Non	Non	
2	LPNC	3	Oui	Oui	Non	Non	Non	Oui	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	

**Figure 3 :** Distribution thématique des différentes (sous-)disciplines étudiées au sein des laboratoires. Cette synthèse a été réalisée sur la base d'un questionnaire de consultation sur le positionnement thématique des unités de recherche rattachées principalement (1 dans la colonne RATT) ou secondairement (2 dans la colonne RATT) à la section 34, en septembre 2023 (laboratoires ayant répondu au questionnaire).

Legendre : PHONET : phonétique; PHONO : phonologie; MORPHO : morphologie; LEX : lexicologie, lexicographie; SYNT : syntaxe; SEMANT : sémantique; PRAG : pragmatique; DISCOURS : discours, texte, dialogue; PHILO : philosophie du langage; EVO : évolution du langage; TYPO : typologie et diversités des langues; HIST-COMP : linguistique historique et comparative; SOCIO : sociolinguistique, contact de langues et dialectologie; PSY-NEURO : psycholinguistique et neurolinguistique; ACQUIS : acquisition du langage; TAL : traitement automatique des langues et linguistique informatique; CORPUS : linguistique de corpus; HIST-LING : histoire de la linguistique. (NB : certaines unités non identifiées comme rattachées secondairement à la section 34 sur l'annuaire du CNRS en 2024, mais qui l'avaient été auparavant, ont aussi répondu au questionnaire de consultation et apparaissent donc ici; EA : laboratoire Éco-anthropologie; GREYC : Groupe de recherche en informatique, image, automatique et instrumentation de Caen; IHRIM : Institut d'histoire des représentations et des idées dans les modernités; LPC : Laboratoire de psychologie cognitive, ayant cofondé au 01/01/2024 le Centre de recherche en psychologie et neurosciences - CRPN).

Enfin, l'annexe 4 résume les autres instruments de recherche dans lesquels la section 34 est impliquée, de manière active et principale, ou bien secondaire.

## B. La démographie des UMR de la section 34

Nous tentons ici de donner un aperçu général de la démographie des chercheurs et chercheuses CNRS, enseignantes-chercheuses et enseignants-chercheurs, ainsi que personnels

ITA CNRS et BIATSS universitaires affectés dans les UMR de la section 34 (cf. annexe 2). Ces données sont issues des bases Réséda de chaque laboratoire, qui ne reflète pas complètement l'actualité au moment de la lecture de ce rapport.

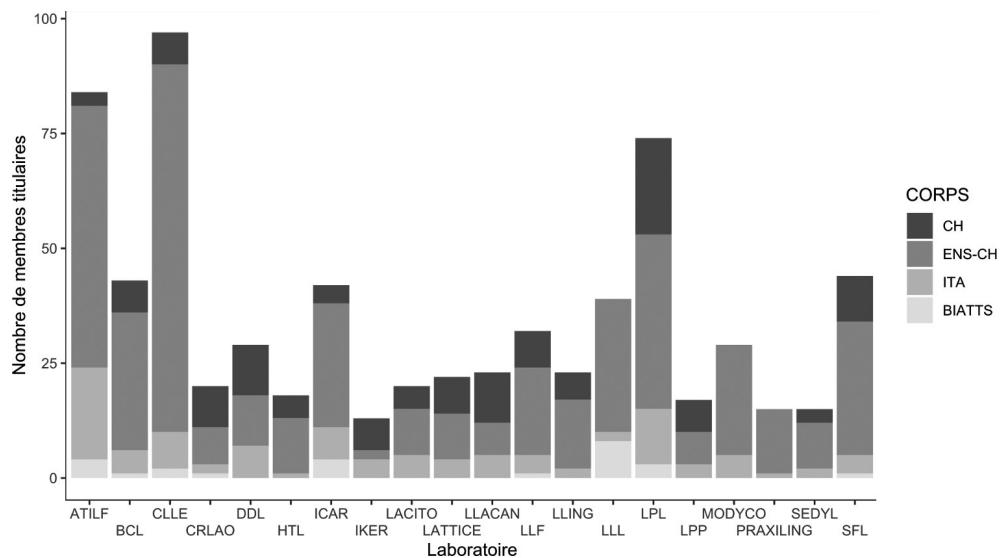
### 1. De manière générale

La figure 4 permet d'apprécier une vue globale de la configuration de chacun des laboratoires. Un point commun à toutes les unités : les personnels universitaires sont plus nombreux que leurs collègues chercheurs et

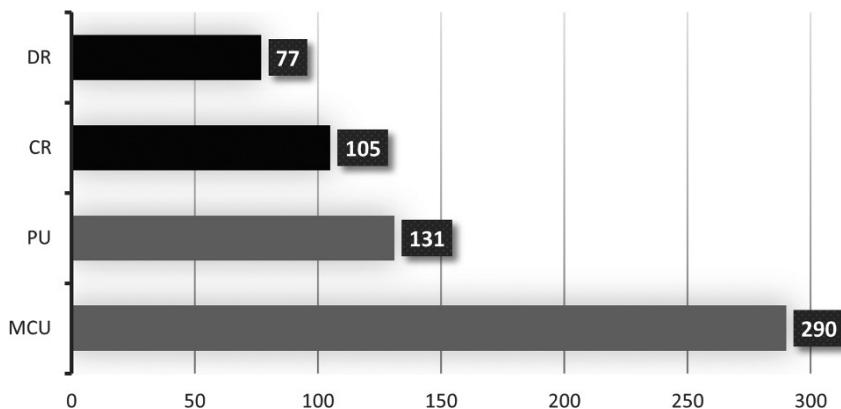
## Comité national de la recherche scientifique

chercheuses CNRS, à quelques exceptions près, qui tendent vers un équilibre entre les deux corps. La figure 5 précise le nombre de personnels dans les deux corps, séparant également les personnels de rang A (professeures/professeurs des universités et directeurs/directrices de

recherche) et de rang B (maîtres/maitresses de conférences et chargées/chargés de recherche) : on note 1,7 PU pour 1 DR, et presque 2,8 MCU pour 1 CR. Concernant les personnels de soutien à la recherche, leurs proportions demeurent assez faibles dans les unités.



**Figure 4 :** Nombre de membres titulaires dans les laboratoires rattachés de manière principale à la section 34.  
CH : chercheurs et chercheuses CNRS ; ENS-CH : enseignantes-chercheuses et enseignants-chercheurs universitaires ; ITA : ingénieries et ingénieres, techniciennes et techniciens et personnels administratifs CNRS ; BIATTS : personnels de bibliothèque, ingénieries et ingénieres, administratifs et administratives, techniques, sociaux et de santé universitaires.



**Figure 5 :** Répartition par catégories des personnels de rang A et B affectés dans les laboratoires rattachés de manière principale à la section 34.  
En noir, chercheurs et chercheuses CNRS ; en gris, enseignantes-chercheuses et enseignants-chercheurs universitaires.

## 2. Le cas spécifique des chercheurs et chercheuses et des ITA CNRS

Le Comité national de la recherche scientifique a une mission spécifique qui concerne l'évaluation, la promotion et le recrutement des chercheurs et chercheuses du CNRS. Il nous a semblé important et cohérent de nous attarder plus spécifiquement sur ces personnels, rattachés à la section 34. Tout d'abord, en ce qui concerne les chercheurs et chercheuses CNRS : pour rappel, nous parlons de 182 personnes (cf. tableau 1 et figure 5), 105 chargées et chargés de recherche et 77 directeurs et directrices de recherche. Au

moment de la rédaction de ce rapport de conjoncture (début 2024), le nombre de chercheurs et chercheuses CNRS rattachés à la section 34 est de **182** (cf. tableau 1 pour une répartition au sein des différents instituts du CNRS). Près de 82% d'entre eux exercent leur activité dans des laboratoires CNRS Sciences humaines & sociales, et 70% dans des laboratoires rattachés à la section 34. Certains chercheurs et chercheuses de la section 34 sont affectés à des laboratoires rattachés à d'autres instituts : CNRS Sciences informatiques (8,8%), CNRS Biologie (7,1%) et CNRS Écologie & Environnement (1,1%).

**Tableau 1 :** Nombre et répartition des chercheurs et chercheuses au sein des instituts du CNRS.

Instituts	Sections pilotes du laboratoire de rattachement	Nombre de chercheurs et chercheuses CNRS rattachés à la section 34
CNRS Sciences informatiques	6 ou 7	16
CNRS Biologie	26	13
CNRS Écologie & Environnement	29 ou 31	2
CNRS Sciences humaines & sociales	32	3
CNRS Sciences humaines & sociales	33	1
<b>CNRS Sciences humaines &amp; sociales</b>	<b>34</b>	<b>127</b>
CNRS Sciences humaines & sociales	35	15
CNRS Sciences humaines & sociales	36	1
CNRS Sciences humaines & sociales	37	1
CNRS Sciences humaines & sociales	38	1
Coévaluation sections 34 et 26		2
	<b>Total de chercheurs et chercheuses en activité</b>	<b>182</b>
	en disponibilité ou détachement	11
	Total	193

La distribution de ces personnels en fonction des corps et grades est évoquée en figure 6. Le nombre d'hommes (n = 83) et de femmes (n = 99) est relativement équilibré. Si la répartition homme/femme est paritaire pour les grades les plus hauts (DR1 et DRCE), les femmes DR2 semblent plus nombreuses, tout comme les CRHC et CRCN (cf. figure 6). Ce nombre plus

élevé de femmes ne se reflète malheureusement pas dans les grades les plus hauts, du fait peut-être du trop petit nombre de possibilités de promotion DR1 et DRCE. Une autre analyse possible serait la tendance générale des femmes à moins candidater à des promotions que les hommes, ce qui est apparent dans le tableau 2 de la campagne d'avancement 2022. Ce tableau

## Comité national de la recherche scientifique

montre que, par exemple, à nombre de promouvables équivalent entre hommes et femmes, l'effectif des candidates aux promotions est

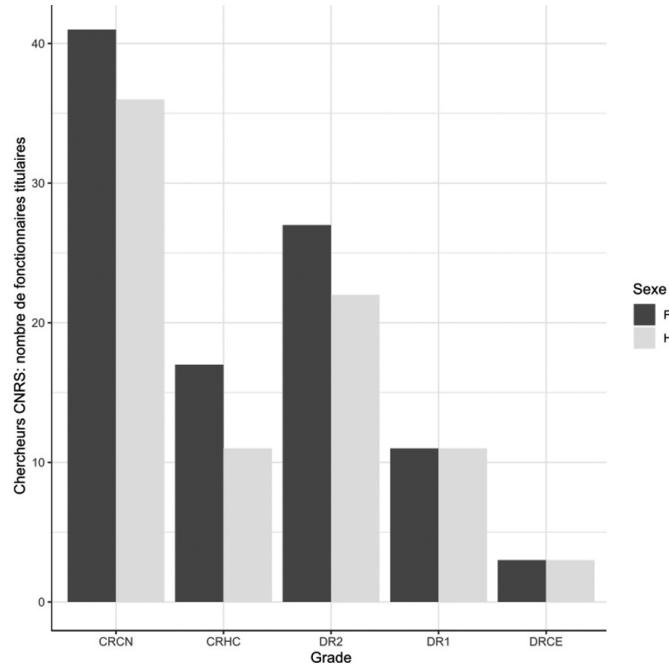
moindre que celui des candidats ; on peut s'interroger sur une potentielle autocensure des femmes dans le cadre de promotions.

**Tableau 2 :** Promouvables hommes/femmes et candidats hommes/femmes pour la campagne d'avancement 2022.

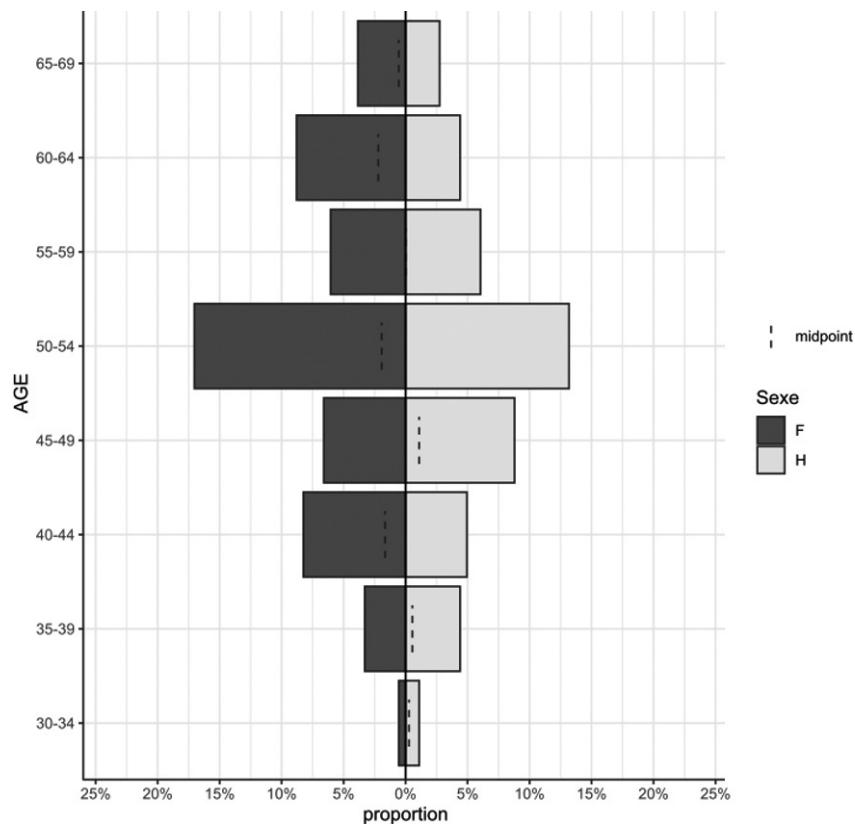
Promotion	Promouvables F	Promouvables H	Candidates F	Candidats H
CRHC	36	32	1	7
CRHC HEB	2	1	2	1
DR1	15	16	5	9
DRCE1	9	6	3	4
DRCE2	1	–	1	–

La pyramide des âges (cf. figure 7) montre une catégorie de personnels principalement (30%) dans la fleur de l'âge (50-54 ans). La proportion des chercheurs et chercheuses en dessous de 40 ans représente moins de 15% de la totalité, ce qui peut être source d'inquiétude quant à la projection du renouvellement de la catégorie 50-54 ans dans 20 ans : quid

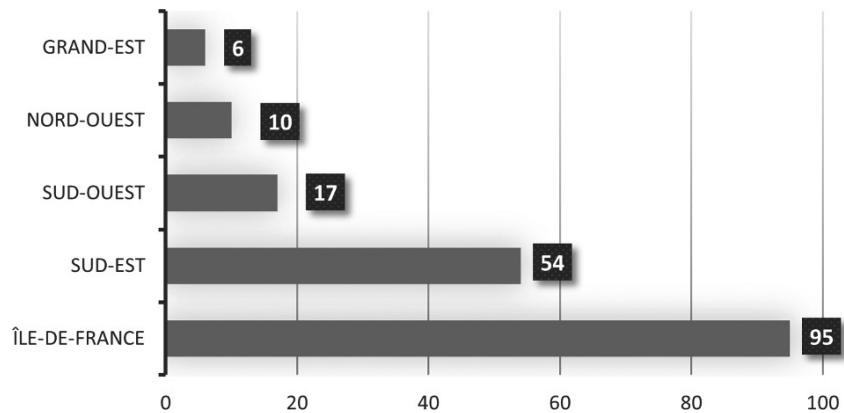
de l'expertise et des responsabilités de fonctionnement des UMR ? Enfin, la distribution des chercheurs et chercheuses dans les différentes régions est corollaire à la répartition des UMR sur le territoire (cf. figure 8) : la moitié est localisée en Île-de-France, tout comme le nombre d'unités (cf. figure 2 et annexe 2).



**Figure 6 :** Nombre de chercheurs et chercheuses CNRS rattachés à la section 34 en fonction de leurs corps et grade, ainsi que la répartition homme/femme.



**Figure 7 :** Pyramide des âges des chercheurs et chercheuses CNRS rattachés à la section 34.

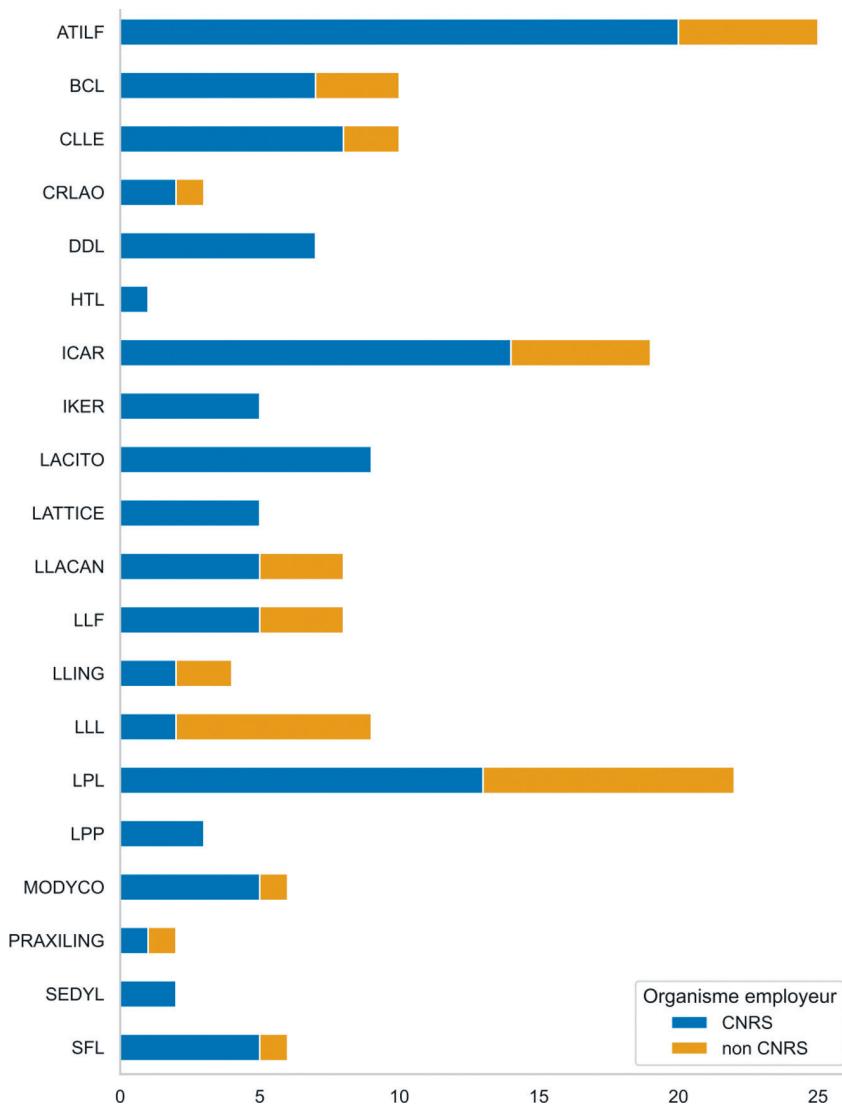


**Figure 8 :** Répartition régionale des chercheurs et chercheuses CNRS de la section 34.

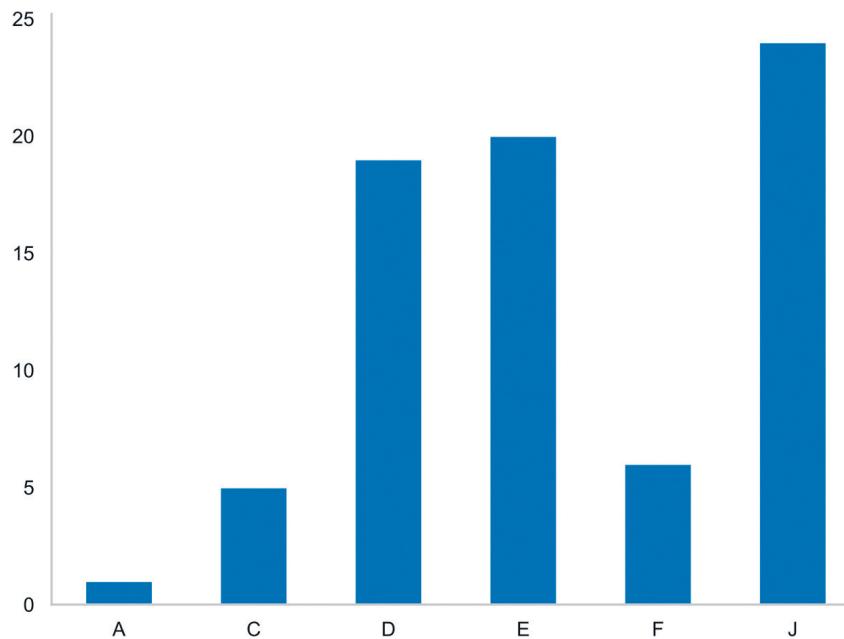
## Comité national de la recherche scientifique

Pour ce qui est du corps des ITA, la situation en fonction des laboratoires est très inégale (cf. figure 9) : certaines unités n'en ont pas d'affectés (certaines n'ont pas non plus de BIATSS), et pour l'ensemble, le nombre et la proportion sont en deçà des besoins. Les laboratoires avec les personnels les plus nombreux ont une répartition de leurs ITA qui semble aux premiers abords en

quantité adéquate, mais ceci n'est pas une généralité. En fonction de la branche d'activité professionnelle (BAP), la distribution des profils de métiers des ITA est représentative des besoins des UMR (cf. figure 10), les pôles représentant le plus de personnel étant Gestion et pilotage (BAP J), Informatique, statistiques et calcul scientifique (BAP E) et Sciences humaines et sociales (BAP D).



**Figure 9 :** Répartition des ITA et BIATSS en fonction de leur laboratoire d'affectation.



**Figure 10 :** Nombre de personnels ITA CNRS dans les UMR rattachées principalement à la section 34 en fonction de leur branche d'activité professionnelle (BAP).

BAP A : Sciences du vivant, de la terre et de l'environnement ; BAP C : Sciences de l'ingénieur et instrumentation scientifique ; BAP D : Sciences humaines et sociales ; BAP E : Informatique, statistiques et calcul scientifique ; BAP F : Culture, communication, production et diffusion des savoirs ; BAP J : Gestion et pilotage.

## II. Les concours de recrutement au CNRS entre 2019 et 2023

La précédente mandature (2016-2021) de la section 34 avait noté une progression nette de la qualité des dossiers des candidats qui se présentent aux concours (et des lauréats de ces concours), notamment au niveau CR, dont la production scientifique et les publications sont souvent d'un niveau exceptionnel. Ils rapportaient également que cette progression concernait globalement toutes les spécialités des sciences du langage, permettant de maintenir une diversité thématique importante dans les recrutements. Notre mandature a réalisé, au moment de la rédaction de ce rapport, deux concours (2022 et 2023), et s'apprête à participer à celui de 2024. Nous ne pouvons

qu'abonder dans le sens de ces constats, et regrettons toutefois que le nombre de postes ouverts au concours annuellement ( $n = 3$  pour 2022, 2023 et 2024) ne permettent malheureusement pas de renouveler suffisamment la dynamique d'un plus large ensemble des domaines de recherche. Comme nos prédecesseurs, nous constatons toutefois une bonne adéquation entre les principes qui guident les travaux de la section et les priorités scientifiques du CNRS, et en particulier du CNRS Sciences humaines & sociales, exprimées notamment dans les coloriages des postes ouverts au concours et dans la mise en place de possibilités de communication et diffusion des sciences du langage auprès des communautés scientifiques élargies. Un autre point important est que la section 34 recrute et promeut très régulièrement des chercheurs et chercheuses qui sont affectés dans des unités d'autres instituts (cf. annexe 3 et tableau 2), et

cela avant même que certains de ces recrutements ne soient institutionnalisés par des postes spécifiques. Ceci est l'exemple fort d'une section tournée résolument vers la pluri- et l'interdisciplinarité (cf. partie III) de son champ de recherche.

## A. Recrutements entre 2019 et 2023

Le tableau 3 répertorie le nombre de postes au concours CR CNRS de 2019 à 2023, et les éventuels coloriages de ces postes. La démographie (H/F et âge) des lauréates et lauréats

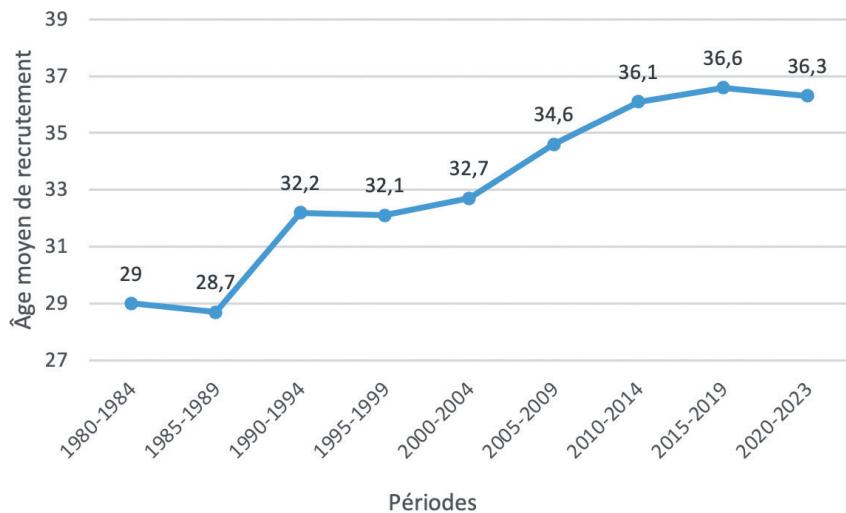
est également mentionnée. Durant cette période de 5 ans, viennent également se rajouter le recrutement d'un CR *via* la CID 53, ainsi qu'un autre par le biais de voie handicap. Ainsi, entre 2019 et 2023, 20 chercheurs et chercheuses sont venus renforcer les effectifs de la section 34. La section 34 se distingue au sein du CNRS par une parité femme-homme satisfaisante dans les recrutements. C'est un important point de vigilance. Elle se distingue aussi par une internationalisation très grande des candidatures aux concours, avec un pourcentage non négligeable de candidates et candidats étrangers, ou diplômés à l'étranger, témoignant du rayonnement et de la forte attractivité des laboratoires de sciences du langage de France.

**Tableau 3 :** Recrutement de chercheurs et chercheuses CRCN de 2019 à 2023.

Année	Nombre de postes au concours	Âge moyen des lauréates et lauréats (min-max)	F/H	Coloriages et précisions
2019	5	33,4 (31-38)	2F/3H	1 poste prioritairement sur le thème : « <i>Morphologie aux interfaces : nouvelles approches</i> » + 1 poste qui développera son projet de recherche dans une unité rattachée à CNRS Biologie à titre principal
2020	3	35 (27-39)	2F/1H	1 poste prioritairement sur le thème : « <i>Intelligence artificielle : linguistique fondamentale et traitement automatique des langues naturelles</i> »
2021	4	37,8 (33-46)	2F/3H	1 poste prioritairement sur le thème : « <i>Troubles de la parole, prosodie et multilinguisme</i> » + <b>1 poste via recrutement CID 53</b>
2022	3	38 (35-44)	1F/2H	1 poste prioritairement sur le thème : « <i>Acquisition typique et atypique du langage</i> »
2023	3	34 (33-35)	2F/2H	1 poste prioritairement sur le thème : « <i>Acquisition typique et atypique du langage et apprentissage</i> » + <b>1 poste voie handicap</b>
2024	3	36 (32-41)	2F/1H	1 poste prioritairement sur le thème : « <i>Sociolinguistique : approches expérimentales et computationnelles</i> »

Afin de continuer l'exploration démographique des recrutements, nous avons complété les données issues du précédent rapport de

conjoncture (cf. figure 11). L'âge moyen de recrutement continue d'avoiser les 36 ans sur la période 2020-2023.



**Figure 11 :** Évolution de l'âge moyen de recrutement des chercheurs et chercheuses CNRS.

## B. Quelques mots sur le concours DR2

Entre 2019 et 2023, le nombre de postes DR2 a sensiblement augmenté d'année en

année, ce qui a été très apprécié à la fois par le comité et par nos collègues (cf. tableau 4). Ceci a aussi été permis, en 2021 et 2023, par le recrutement de DR2 « externes » supplémentaires, du fait de candidatures de qualité scientifique exceptionnelle.

**Tableau 4 :** Promotions DR2 de 2019 à 2023.

Année	Nombre de postes au concours	Âge moyen des lauréates et lauréats (min-max)	F/H	Précisions
2019	5	48,2 (39-59)	3F/2H	
2020	5	45,6 (39-57)	4F/1H	
2021	4	43,6 (39-48)	2F/3H	<b>+ 1 lauréat externe</b>
2022	4	45,3 (41-48)	2F/2H	
2023	6	51,4 (44-56)	4F/3H	<b>+ 1 lauréate externe</b>
2024	7	49,8 (42-63)	4F/4H	<b>+ 1 lauréat externe</b>

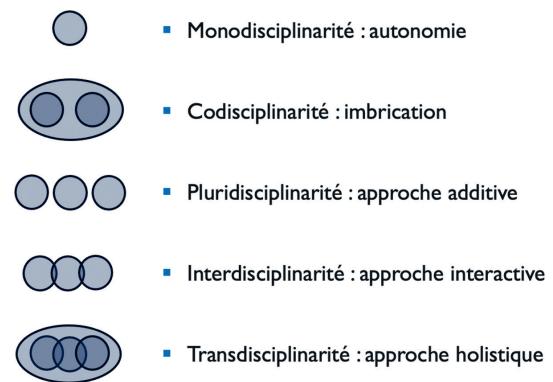
### III. L'interdisciplinarité des sciences du langage

Dans le domaine des sciences du langage, l'interdisciplinarité a connu une évolution significative depuis les années 1980. Les approches disciplinaires, puis pluridisciplinaires, ont progressivement laissé place à des approches interdisciplinaires, voire même transdisciplinaires, qui ont intégré le périmètre des sciences du langage au fil du temps, démontrant ainsi une dynamique complexe et mouvante, à la fois au niveau des objets et des méthodes déployées. Cette évolution témoigne d'une reconnaissance croissante de l'importance du dialogue et des échanges entre disciplines des sciences du langage et au-delà, favorisant une approche plus holistique et complète de l'étude des langues et du langage dans leurs multiples dimensions sociales, culturelles, cognitives et technologiques. Cette évolution est également visible dans nos laboratoires, où le paysage est de plus en plus interdisciplinaire : les thématiques de recherche, les projets proposés et certains de nos recrutements montrent l'ambition pour des approches de plus en plus transversales.

#### A. Quelques définitions

Ce que recouvre réellement le terme « interdisciplinarité » dans notre communauté, où se confondent souvent des concepts aussi complexes que la codisciplinarité, la pluridisciplinarité, l'interdisciplinarité et la transdisciplinarité, n'est pas complètement évident (cf. figure 11). Ces termes témoignent indiscutablement d'une diversité d'approches et de méthodes employées, non seulement en sciences du langage mais aussi au-delà. « L'interdisciplinarité » fait typiquement référence à l'interaction, au travail en équipe, à la capacité des chercheurs issus de différentes disciplines de déployer ou de développer des outils, des méthodes ou des approches conceptuelles variées pour apporter des éléments de réponse

à des défis scientifiques communs et complexes. Mais cette « interdisciplinarité » peut prendre des formes variées, telles que théorisées par exemple, par Choi & Pak, 2006 (cf. figure 11), et questionne la place du travail disciplinaire en son sein.



**Figure 12 :** Représentation schématique du travail mono-, co-, pluri-, inter- et transdisciplinaire (inspirée par Choi & Pak, 2006).

La **monodisciplinarité** prône l'autonomie, à la fois dans la définition des objets et dans la conception des outils de travail. Quant à la **codisciplinarité**, elle concerne l'étude d'un objet à partir de plusieurs disciplines imbriquées l'une dans l'autre et indissociables. Par exemple, on ne peut pas étudier l'évolution des langues et les parentés entre systèmes, l'une des branches les plus anciennes de la linguistique, sans y associer la géographie, ou encore la paléontologie ou l'anthropologie.

La **multi- ou pluridisciplinarité** implique des approches multiples qui consistent à envisager un objet d'étude par la juxtaposition de points de vue, souvent sans volonté de dégager un lien ou d'apporter une réponse unique et uniforme. L'objet d'étude est vu à travers différentes dimensions. C'est le cas, par exemple, des études sur corpus ou du travail des typologues souvent amenés à proposer des descriptions ou des annotations multicouches (phonologiques, sémantiques, syntaxiques, etc.) de manière indépendante pour des exploitations, caractérisations et classifications objectives des phénomènes observés.

Au niveau de l'**interdisciplinarité** à proprement parler, le dialogue entre disciplines devient plus fonctionnel et sans frontières étanches. Il s'agit, à partir d'une discipline, de se demander et de voir ce que les disciplines connexes apportent de plus en termes de connaissance. Le résultat est une interaction, un croisement, à la fois au niveau des démarches méthodologiques adoptées et des résultats observés, apportant un enrichissement, et par conséquent une compréhension plus globale, voire systémique, de l'objet étudié. C'est le cas, par exemple, des chercheurs qui travaillent à l'interface phonologie-syntaxe qui étudient la production langagière de manière parallèle, à la fois sous ses aspects structurels et formels, ou le cas des chercheurs qui travaillent à l'interface sémantique-pragmatique et étudient la dynamique du sens à différents niveaux et contextes d'analyse. Par exemple, la psycholinguistique ou la neurolinguistique recourent systématiquement aux méthodes comportementales, psychophysiques, électrophysiologiques, ou d'imagerie cérébrale pour apporter des réponses et une compréhension multidimensionnelle du comportement langagier et/ou de ses pathologies. Cela peut également conduire à des solutions plus pratiques par exemple grâce à la modélisation computationnelle, ou aux applications cliniques de réhabilitation du langage.

Quant à la **transdisciplinarité**, elle se nourrit fortement des approches interdisciplinaires, mais son objectif principal est de traverser les approches, résultats, points de vue disciplinaires et de dégager de manière plus holistique des éléments transversaux qui informent de manière bidirectionnelle, voire multidirectionnelle, les différentes disciplines impliquées. Cette approche va au-delà des champs disciplinaires, et considère l'objet d'étude en tant que système en l'envisageant dans sa complexité, et qui surtout adopte une posture scientifique, épistémologique et intellectuelle qui se situe à la fois entre, à travers et au-delà des approches compartimentées. Un programme transdisciplinaire de ce type serait, par exemple, de combiner acquisition du langage et linguistique computationnelle comparant les trajectoires d'apprentissage

potentielles des modèles de langage génératifs de l'IA avec les processus d'acquisition du langage chez les enfants dans un but double : d'informer les modèles du développement du langage, et de rendre les algorithmes d'entraînement plus efficaces en les entraînant à l'échelle humaine, avec des données naturalistes et, idéalement, multilingues.

Les approches interdisciplinaires et transdisciplinaires émergent dans un paysage où l'on constate une complexité croissante du monde, un éclatement de la connaissance, une pluralité des savoirs, et surtout une nécessité de décloisonnement pour mieux comprendre la nature et l'articulation du paysage dans lequel évolue l'être humain. Les pratiques inter- et transdisciplinaires s'opèrent selon deux modes qui reposent : (i) sur une **complémentarité** entre compétences disciplinaires ; et (ii) sur la capacité d'**adaptation** et d'emprunt des concepts, modèles, mesures ou outils d'observation issus d'autres disciplines. Cette double articulation se fait au service d'un objectif commun – par exemple, dans le domaine des sciences du langage, pour étudier l'être humain, ses capacités langagières, ses systèmes symboliques, permettant ainsi de s'ouvrir vers des nouveaux objets, des nouvelles méthodes, des simulations ou des modélisations.

De telles approches engendrent inévitablement des défis, dus par exemple à la diversité des pratiques, la variété des objets et des méthodes employées, ce qui amène souvent les scientifiques à des rivalités entre ceux qui défendent le noyau dur d'une discipline et ceux qui proposent des études plus holistiques et un décloisonnement des approches. Le décloisonnement proposé peut venir de différentes manières en sciences du langage : à travers des pratiques inter- et transdisciplinaires, mais également à travers des initiatives d'interfaces disciplinaires ou codisciplinaires.

Au sein des sciences du langage, nous pouvons distinguer au moins trois types d'**interdisciplinarité** : celle issue d'interrelations à l'intérieur même de nos disciplines (par exemple, phonologie-syntaxe ; sémantique-pragmatique) ; celle qui nécessite une interaction avec des disciplines externes comme

l'histoire, l'anthropologie, la philosophie, la psychologie, la littérature, l'informatique, etc. ; et celle qui fait interagir les sciences du langage avec des groupes de sciences comme les sciences de l'éducation ou les sciences cognitives.

## B. L'interdisciplinarité au CNRS

Du point de vue institutionnel, CNRS Sciences humaines & sociale développe un cadre afin de mener ce type de recherche dans une perspective internationale et interdisciplinaire au sens large. L'Institut soutient de différentes manières ce type d'initiatives, par exemple, à travers **les actions de la DAS interdisciplinarité** et de la **MITI** (Mission pour les initiatives transverses et interdisciplinaires), qui ont pour objectif principal de soutenir des programmes de recherche interdisciplinaires et des initiatives transverses au sein du CNRS, soutenir les prises de risques, la rupture et l'innovation.

Au niveau du CoNRS, **les commissions interdisciplinaires** (CID), telles que la CID 51 « Modélisation mathématique, informatique et physique pour les sciences du vivant », la CID 52 « Environnements sociétés : du savoir à l'action », la CID 53 « Sciences en société : production, circulation et usages des savoirs et des technologies » et la CID 55 « Sciences et données » ont également un rôle très important dans la promotion de l'interdisciplinarité, en particulier à travers le recrutement de jeunes chercheurs qui proposent ce type d'ouverture. Tout cela est complété par **le travail effectué dans les sections**, comme la nôtre (section 34), à travers les coloriages des postes proposés chaque année (cf. tableau 3), souvent à l'interface entre linguistique, informatique, sciences cognitives, sciences sociales, psychologie, philosophie, etc., et à travers les financements accordés à des groupements et réseaux de recherche initiés par nos chercheurs. À titre d'exemple, nous pouvons mentionner cinq **initiatives fédératrices et interdisciplinaires soutenues par la section 34** ces dernières années (cf. annexe 4) :

- l'infrastructure de recherche Humanités numériques (Huma-Num) ;
- le réseau thématique pluridisciplinaire (RTP) Éducation ;
- le réseau Linguistique informatique, formelle et de terrain (LIFT, LIFT2) ;
- le réseau thématique Acquisition des langues secondes (Real, ReaL2) ;
- le réseau Langues et langage à la croisée des disciplines (LLcD).

La section 34 soutient aussi chaque année des projets IRN (International Research Network) inter-/transdisciplinaires, comme par exemple le projet OASIS (*Ontology as Structured by the Interfaces with Semantics*) qui inclut des linguistes théoriques en sémantique et syntaxe, des psychologues cognitivistes, des philosophes, et des linguistes computationnels.

## IV. État des lieux scientifique des sciences du langage

### A. Linguistique fondamentale

#### 1. Phonétique, phonologie

La phonétique vise à décrire et à analyser les mécanismes de production et de perception de la parole, ainsi que le produit, acoustique ou perceptif, de ces mécanismes. Elle recouvre aussi bien la description que la modélisation de la production et la perception de la parole du point de vue acoustique et articulatoire (par exemple, relations entre perception et production, calcul de l'articulation à partir du signal acoustique). On observe une tendance forte à travailler à l'interface entre phonétique et phonologie de laboratoire et psycholinguistique notamment sur l'acquisition et le bilinguisme. D'autres axes de recherche interdisciplinaires concernent la description des systèmes sonores des langues en lien avec la phonologie et la typologie, l'étude diachronique des langues,

les variétés régionales et les dialectes, champs qui se révèlent de plus en plus productifs. La phonétique clinique, et notamment les travaux sur les pathologies de la parole, constitue un autre secteur particulièrement important de domaine.

La plupart des recherches en phonétique est de type expérimental (en production : analyses acoustiques et analyse des gestes des articulateurs ; en perception : tests comportementaux de discrimination, de catégorisation, etc., électrophysiologie, imagerie cérébrale). Le recueil et l'analyse de corpus, allant du contrôlé au spontané, sont largement représentés, avec, pour l'acquisition de la parole, des corpus de parole enfantine ou de parole adressée aux enfants dans des contextes d'interaction. La modélisation acoustique et articulatoire est également présente dans les études de production bien que les recherches sur les modèles articulatoires devraient être développées davantage.

Parmi les domaines interdisciplinaires auxquels la phonétique apporte une contribution, la sociophonétique et la neurophonétique apparaissent comme particulièrement productives. Le premier domaine croise les approches théoriques et les méthodes propres à la phonétique expérimentale avec celles des études variationnistes venant notamment de la sociolinguistique et de la psycholinguistique (cf. parties IV.B.2 et IV.B.3) ; dans ce cadre, les nouvelles études sur la voix et le genre prennent de l'importance. Le deuxième domaine, à l'interface entre phonétique, neurologie et sciences cognitives, s'intéresse à l'acquisition et aux pathologies de la parole avec une attention grandissante pour la parole multimodale. Notons enfin que l'apport des études phonétiques aux études sur la prosodie des langues, y compris dans le cadre du contact linguistique tend à se renforcer.

La phonologie se consacre à l'identification et à l'étude des unités minimales cognitives du langage oral, ainsi que des principes qui régissent leur distribution et leur fonction. Si phonétiquement on observe une forte variation des réalisations sonores (acoustiques, articulatoires, gestuelles), la phonologie examine les

représentations permettant de ramener cette diversité à une identité sous-jacente, à un nombre restreint de catégories (phonèmes, syllabes, gabarits, configurations manuelles en langue des signes) et de principes. Elle s'attache à définir le contenu intrinsèque de ces catégories (traits binaires/unaires, constituants syllabiques, tons), les processus qui les affectent (harmonie, allongement, assimilation, dissimilation, métathèse, liaison, activité gabaritique) et qui ne peuvent s'expliquer sur la seule base de la substance. Par sa dimension symbolique, la phonologie entretient naturellement des liens avec la morphologie et la syntaxe, disciplines avec lesquelles elle partage une partie de ses outils d'analyse et de formalisation.

La majeure partie des recherches en phonologie en France s'inscrit dans des modèles héritiers du cadre formel génératif : phonologie du gouvernement et théorie de l'optimalité. Si l'on observe une proximité naturelle avec la phonétique, et notamment avec la phonologie de laboratoire, l'émergence actuelle des approches phonologiques des langues signées ne doit pas être négligée. Notons également l'importance croissante des travaux basés sur corpus et le développement des approches quantitatives et/ou mixtes qui remettent en question la dichotomie phonétique/phonologie, voire qui la dépassent.

Les recherches à l'interface de la phonologie et de la phonétique et/ou de la psycholinguistique visent à mieux comprendre et définir le lien entre l'aspect cognitif et conceptuel de la parole humaine et son aspect physique, et traitent principalement de questions liées à la nature des représentations phonologiques chez l'adulte, l'enfant ou le nourrisson. La démarche expérimentale en phonologie est fortement visible dans les travaux employant les méthodes de la psychologie expérimentale et dans les recherches fondées sur des approches computationnelles qui visent à proposer des modélisations plausibles du point de vue cognitif dans le cadre des études sur l'acquisition et l'apprentissage des catégories phonémiques.

Enfin, l'éventail de langues traitées dans le cadre de la phonologie est large et s'appuie sur des données issues d'un travail de terrain permettant d'asseoir les analyses sur une base empirique fiable.

## 2. Morphologie

La morphologie est l'étude de la forme interne des mots et des éléments qui les composent, par opposition à la syntaxe qui étudie les combinaisons de mots entre eux. Elle couvre plusieurs sous-domaines : la morphologie flexionnelle, qui s'intéresse aux paradigmes de flexion d'un lexème (par exemple, marquage de personne, temps/aspect, pluriété, définitude, etc.) et la morphologie dérivationnelle ou constructionnelle, qui s'intéresse à la formation des lexèmes. Selon que l'accent est mis sur la forme, le sens, ou sur les conditions d'utilisation, d'émergence ou de mise en mémoire des mots, les recherches s'inscriront également dans le champ de la phonologie, de la sémantique, de la syntaxe, de la pragmatique ou de la psycholinguistique. La morphologie se prête bien à la modélisation cognitive et informatique. Elle donne lieu à des travaux relevant du traitement automatique des langues et de la création de bases de données et de ressources.

Toutes ces orientations se retrouvent dans l'éventail des recherches actuellement menées en France, dans des cadres théoriques variés, qui vont de la morphologie paradigmique, en passant par la morphologie lexématique jusqu'à la morphologie dite distribuée. Les méthodologies sur lesquelles ces travaux se fondent tendent à s'appuyer sur des corpus de grande taille et relèvent pour certains de la linguistique expérimentale.

Une nouvelle orientation prometteuse pour l'étude de l'organisation morphologique du lexique est fondée sur l'exploitation de ressources lexicales participatives comme les dictionnaires collaboratifs. Ces travaux se situent à l'interface entre morphologie, linguistique de corpus et traitement automatique des langues.

## 3. Syntaxe

La syntaxe étudie la façon dont les unités de signification (morphèmes, mots) se combinent entre elles pour engendrer des unités de signification plus grandes (syntagmes, phrases simples et complexes). Elle cherche à déterminer les principes qui gouvernent l'organisation des combinaisons possibles et impossibles d'unités dans les langues naturelles, en lien avec les mécanismes combinatoires de leur interprétation.

Les recherches en syntaxe sont diversifiées, tant du point de vue des paradigmes scientifiques et méthodologiques représentés, qu'au niveau des aires linguistiques couvertes. Les recherches sont conduites à la fois dans un but descriptif (description explicite de la grammaire de différentes langues vivantes ou anciennes parlées dans le monde, du point de vue synchronique ou diachronique) et/ou dans un but théorique (par exemple, la mise au jour de paramètres, de principes et contraintes universels, et principes de micro-/macrovariation entre langues). Les cadres théoriques adoptés incluent les approches génératives (représentées principalement par le programme minimaliste, HPSG et LFG) et les approches fonctionnalistes (par exemple, RRG, *Functional Grammar*, approche fonctionnelle évolutive, etc.). Une large partie des recherches en syntaxe porte sur l'interface avec les autres niveaux d'analyse, notamment la sémantique, la morphologie, la phonologie et la prosodie.

La recherche en syntaxe se définit aussi par l'ouverture vers d'autres disciplines, y compris le traitement automatique des langues (TAL), la psychologie développementale, les neurosciences, la philosophie du langage, les sociologie et anthropologie culturelles. Ces approches interdisciplinaires comportent une forte dimension expérimentale qui met les hypothèses de la syntaxe théorique à l'épreuve des données de terrain, de traitement, de compréhension ou de production, en passant souvent par des études de corpus et en recourant à des méthodes d'analyse quantitative et neurophysiologiques.

Les recherches en syntaxe couvrent l'ensemble des aires linguistiques du monde, sans distinction de leur reconnaissance officielle (tout « dialecte » est une langue), de leur statut comme langue écrite, à tradition orale ou signée. Une attention particulière est portée ces dernières années à l'étude des langues en danger.

#### 4. Sémantique

La sémantique a pour sujet le sens des mots et des constructions linguistiques (syntagmes, phrases, discours). Elle s'intéresse notamment à la manière dont le sens d'une expression complexe dépend du sens de ses éléments plus simples (compositionnalité). Son champ inclut l'étude (i) des catégories lexicales à pertinence grammaticale (sémantique des classes de noms, des classes de verbes et types de situations, des expressions gradables, vagues ou scalaires, etc.) et de l'ontologie qu'elles supposent ; (ii) des items lexicaux appartenant aux catégories fonctionnelles qui constituent « l'appareil quantificationnel du langage » (déterminants, nombre grammatical, temps, aspect, modalité, etc.) ; (iii) des questions d'interface syntaxe-sémantique comme l'interprétation des pluriels, la portée relative des quantificateurs, l'interprétation de l'ellipse et des anaphores et la relation entre la sémantique lexicale et la structure syntaxique ; et (iv) de la part qui revient à l'intégration de diverses dimensions de sens dans le calcul du contenu d'une phrase ou d'un texte comme la présupposition, les inférences pragmatiques, ou la dimension émotive ou évaluative.

Depuis plusieurs décennies, une grande partie des sémanticiens dans le monde modélise le sens en utilisant la théorie des modèles issue de la logique et de la sémantique des langages formels, dans laquelle la notion centrale est celle de conditions de vérité (en première approximation, le sens d'un énoncé est assimilé à l'ensemble des situations possibles dans lesquelles l'énoncé est vrai). Un développement important a consisté à passer des modèles vériconditionnels à des approches dynamiques permettant une intégration de la

sémantique de la phrase et du discours. L'approche vériconditionnelle a également été enrichie pour rendre compte de l'interprétation des énoncés non déclaratifs (interrogatifs, impératifs). D'autres approches sont également utilisées, comme les grammaires cognitives ou la théorie des opérations énonciatives.

On constate dans le paradigme de la sémantique formelle des ouvertures importantes conduisant à l'intégration (i) d'aspects discursifs et dialogiques, en général dans le cadre d'approches dynamiques du sens, avec un intérêt accru pour la modélisation du sens des phrases non déclaratives en contexte, des actes de langage autres que l'assertion, et de marqueurs discursifs et dialogiques ; (ii) de la dimension sociale de la construction de la signification, avec des travaux qui utilisent à la fois des méthodes formelles et celles de la sociolinguistique ; (iii) d'approches empiriques nouvelles, et ceci dans les deux domaines que constituent le travail sur les grands corpus et les expériences psycho-, voire neurolinguistiques ; (iv) de données provenant de domaines nouveaux, en particulier les langues des signes, qui fournissent un terrain particulièrement propice à l'exploration des contraintes sémantiques universelles sur des phénomènes de deixis, d'attitudes propositionnelles, d'ellipse, d'anaphore, des gestes coverbaux ou proverbiaux et, au-delà du langage, de la musique.

La sémantique connaît actuellement un renouveau important, la conduisant à adopter certains des concepts et des méthodes de l'informatique et de la psychologie cognitive. On note aussi le rôle nouveau que prennent les approches distributionnelles fondées sur les méthodes d'apprentissage automatiques, qui s'appuient en particulier sur l'apprentissage profond au moyen de réseaux de neurones. Notons également l'émergence de théories probabilistes de la sémantique et de la pragmatique, modélisant l'interprétation comme un processus de raisonnement sous incertitude *via* des méthodes fréquemment utilisées dans les sciences cognitives et l'intelligence artificielle pour traiter la perception, le raisonnement et la construction des concepts.

## 5. Pragmatique, philosophie du langage

L'interprétation d'une phrase ou d'un discours, dans un contexte donné, dépend à la fois de sa signification linguistique littérale (déterminée par le sens des mots qui y apparaissent et par la structure syntaxique), de certains paramètres du contexte, et des inférences que les interlocuteurs dérivent au sujet de l'état épistémique du locuteur et de ses intentions communicatives. La pragmatique s'intéresse à tous les aspects contextuels de la signification linguistique. Nombre d'expressions et constructions reçoivent leur contenu du contexte (par exemple, les pronoms, les démonstratifs, les prédictats gradables), mais l'importance du contexte dans l'assignation d'un contenu aux expressions est en réalité extrêmement générale. De plus, les énoncés ont non seulement des conditions de vérité, mais aussi des conditions d'usage qui font référence au contexte, conditions qui font souvent partie du sens proprement linguistique des expressions – présuppositions lexicales, notamment. Au-delà, la pragmatique s'intéresse aussi à tous les aspects conventionnels de la signification qui ne se réduisent pas aux conditions de vérité, parfois rangés dans la rubrique des implicatures conventionnelles. D'autre part, la pragmatique s'intéresse à la dimension inférentielle de l'interprétation. Elle cherche à comprendre la manière dont les interlocuteurs enrichissent le sens proprement linguistique pour interpréter les énoncés dans leur contexte en raisonnant sur les intentions des interlocuteurs – phénomènes d'implicatures conversationnelles.

Les recherches en pragmatique se développent selon plusieurs approches : une approche formelle et expérimentale, qui sont souvent combinées, et une approche énonciative. La pragmatique formelle opère généralement de concert avec les travaux en sémantique formelle et en sémantique du discours, et vise à produire des modèles explicites des phénomènes de dépendances contextuelles (pronoms, présuppositions, implicatures conventionnelles), au moyen de formalismes inspirés par la logique, et plus récemment, au moyen de modèles probabilistes bayésiens ou relevant de la théorie des jeux. L'approche

expérimentale relève souvent de travaux de la psychologie cognitive, en particulier de la psychologie du raisonnement et de la cognition sociale, et comprend aussi une dimension développementale (étude du développement des capacités pragmatiques chez les enfants). Une partie des travaux conduits en pragmatique expérimentale vise aussi à tester les prédictions des modèles formels ou à dégager les mécanismes cognitifs sous-jacents aux généralisations dégagées par les linguistes. Les méthodes employées incluent notamment l'électro-encéphalographie (EEG) et les méthodes comportementales. Une dernière direction émergente, relève de travaux empiriques, notamment des approches herméneutiques et praxéologiques de corpus (pragmatique de corpus) qui étudient les dimensions énonciatives, interactionnelles et multimodales du discours.

La philosophie du langage se rattache à la philosophie analytique, à la sémantique et à la logique, tout en ayant une forte préoccupation pour des thématiques pragmatiques et sémantiques, principalement liées à la communication implicite. Parmi les questions linguistiques qui font l'objet de recherches actives en philosophie du langage, on compte l'indexicalité, les attitudes propositionnelles (concepts correspondant aux verbes croire, penser, savoir, vouloir, etc.), l'intentionnalité et l'hyperintentionnalité, le vague, l'interprétation des énoncés conditionnels, la sémantique des termes dépréciatifs comme les insultes et des termes évaluatifs.

L'intérêt pour la communication implicite a contribué de façon cruciale au débat sur le contextualisme qui concerne la détermination de la frontière entre sémantique et pragmatique dans l'interprétation des énoncés et également sur les dimensions expressives (non propositionnelles) de la signification linguistique. Par ailleurs, la philosophie du langage a développé des travaux sur les rapports entre langage et pensée, sur la référence et les fichiers mentaux, sur le rôle de la perspective dans la production des énoncés, ou sur de nouvelles approches du vague articulant modèles psychologiques, logiques et

probabilistes. Elle s'intéresse aussi aux liens entre langage et ontologie (conceptuelle et grammaticale), et à la clarification des concepts, en s'appuyant sur les résultats de la sémantique linguistique.

D'autres approches existent, aux postulats souvent aux antipodes de la philosophie analytique, inspirées par la phénoménologie, le poststructuralisme, la déconstruction, la philosophie de la littérature, la psychanalyse, les féminismes, et qui influencent les réflexions sur la nature instable du sens et la frontière imprécise entre sens littéral et sens dérivé (par exemple métaphorique), l'inscription du langage individuel dans le monde et la société et certaines formes d'analyse du discours.

## 6. Texte et discours

Les recherches sur le discours, les textes et le dialogue concernent les approches linguistiques portant sur la structure et le contenu de productions plus longues qu'une phrase, et incluant souvent multidimensionnalité et multimodalité. Une hypothèse commune aux travaux dans ce domaine est l'idée que ces productions ne sont pas seulement des suites de phrases ou d'énoncés, et que leur contenu n'est pas seulement une conjonction ou une intersection des valeurs sémantiques de ces phrases.

Au sein de l'analyse du discours, la linguistique interactionnelle s'intéresse plus spécifiquement à l'analyse et à la modélisation de processus cognitifs mis en œuvre dans les interactions communicatives finalisées. L'interaction y est conçue comme la forme fondamentale de sociabilité, de contexte de raisonnement pratique ainsi que le lieu d'émergence et de stabilisation de la grammaire. Elle peut dès lors contribuer à l'élaboration d'une démarche théorique et méthodologique apte à rendre compte de l'émergence de la cognition et de la connaissance, dans et par le dialogue. Les interactions analysées sont situées dans des contextes sociaux spécifiques (éducatifs, commerciaux, médicaux, etc.) ou médiées par ordinateur. L'un des enjeux de ces recherches est de documenter la dimension multimodale des

interactions à distance. Avec l'explosion de la société de l'information, l'analyse du discours prend, théoriquement et empiriquement, une nouvelle dimension autour de l'étude des médias sociaux et des usages médiés de la langue, avec de nombreuses ramifications applicatives.

L'analyse de discours produit des analyses formelles de l'interaction dialogique capables de capter de manière précise les aperçus empiriques de la linguistique interactionnelle et de la psychologie cognitive, notamment en collaboration avec le TAL pour créer des systèmes de dialogue parlés comme les assistants vocaux. Les efforts se sont aussi concentrés sur l'interaction entre approches computationnelles (méthodes statistiques d'apprentissage automatique ; méthodes hybrides utilisant à la fois des méthodes symboliques et statistiques) et études formelles de la structure et du contenu du discours (SDRT, RST, DLTAG).

Le développement récent des initiatives de normalisation et d'échange de corpus textuels, oraux et vidéo, des systèmes pour leur annotation linguistique et sémiotique et enfin leur instrumentation informatique, permet à l'analyse du discours de se développer autant dans sa dimension descriptive, comparative et théorique, en lien avec des sous-disciplines des sciences du langage telles que la linguistique cognitive ou la sociolinguistique, et des sous-disciplines en dehors des sciences du langage, en lien avec les sciences cognitives, la psychologie, la théorie littéraire et les sciences sociales pour la modélisation des paramètres sociaux-cognitifs impliqués dans l'acte communicationnel.

## 7. Lexicologie, lexicographie, traductologie

La lexicologie étudie les unités lexicales, les mots et les expressions figées mais également les relations qu'elles entretiennent entre elles en diachronie (étude du lexique dans une perspective historique et philologique ; réalisation de dictionnaires étymologiques) et synchronie (constitution de dictionnaires ; analyse des interactions entre lexique et syntaxe, entre

lexique et sémantique ; évolution du lexique dans sa mise en discours ; place du lexique dans la cognition). Les recherches lexicales, terminologiques et phraséologiques se servent souvent des méthodes de la linguistique de corpus pour l'analyse des collocations, de la néologie ou encore des langues de spécialité. Toujours en lien avec la linguistique de corpus, la lexicométrie s'intéresse à la fréquence et aux cooccurrences des unités lexicales dans des corpus textuels, ainsi qu'à leur sens et leur usage.

Les travaux de lexicologie et lexicographie, et notamment la constitution de dictionnaires et de grands corpus, couvrent un grand nombre de langues bien étudiées et/ou sous-étudiées et en danger.

Un développement récent en lexicographie est la création participative de dictionnaires de qualité. Cette nouvelle lexicographie collaborative et les ressources lexicales qui en dérivent restent peu explorées dans les travaux plus traditionnels et sont principalement exploitées par d'autres secteurs de la linguistique comme la morphologie (cf. partie IV.A.2), la linguistique de corpus (cf. partie IV.B.1), la didactique des langues et pour la remédiation auprès de personnes atteintes de pathologies de la parole et de la lecture.

## B. Approches transversales

### 1. Linguistique de corpus, traitement automatique du langage, linguistique informatique

La linguistique de corpus (LC), le traitement automatique des langues (TAL) et la linguistique informatique (LI) reposent sur l'utilisation d'outils informatiques et algorithmiques pour décrire des phénomènes langagiers. La LC se concentre sur des aspects plus descriptifs, quand le TAL et la LI ont des objectifs de modélisation formelle, qui peuvent se traduire par des applications et des méthodes opérationnelles qui peuvent nourrir la LC.

Les linguistes qui mobilisent dans leurs travaux des corpus annotés tirent directement

avantage des avancées du TAL qui se sont concrétisées ces trente dernières années par la mise en place de chaînes d'annotation de plus en plus complètes et performantes, d'abord pour l'étiquetage morpho-syntaxique et la lemmatisation, ensuite pour les annotations de haut niveau, notamment les relations de dépendance syntaxique. Ces enjeux sont de taille, car ils ont trait à la conception des méthodologies de requête et de leur possible simplification par le recours à des annotations pouvant servir, le cas échéant, de repères de synthétisation des jeux de données soumis à l'analyse linguistique. De ce fait, faute de ressources d'annotation stabilisées, les études sur (et par) corpus qui s'inscrivent dans les domaines de la sémantique ou la linguistique textuelle se trouvent toujours face à des verrous considérables en ce qu'elles sont reléguées à des méthodologies de requête multiparcours d'une certaine complexité, les corpus dotés d'annotations sémantiques de haut niveau (par exemple rôles sémantiques ou chaînes de coréférence) étant rares et plutôt de petite taille (on voit bien sûr ici en parallèle tous les défis liés d'une manière générale à la question des langues dites « peu dotées »). Une stratégie de contournement déployée actuellement par certains analystes de corpus consiste à remplacer l'interrogation directe d'étiquettes catégorielles explicites par la mobilisation directe de modèles vectoriels génériques (*embeddings*), quitte à accepter la mise en place de méthodologies de requête (partiellement) opacifiées. Pour ce qui est des méthodologies d'exploration statistique (qui dépendent évidemment de la fiabilité des requêtes d'extraction réalisées en amont), la tendance à l'utilisation de procédés de synthétisation multivariée (*clustering*, échelonnement multidimensionnel, analyses factorielles), qui n'est certes pas tout récente (comme le montre entre autres la vitalité, depuis plus d'une trentaine d'années, de la communauté des recherches en ADT), reste intacte, ce qui est dû à la disponibilité de plus en plus généralisée d'outils informatiques qui proposent un large éventail de méthodes de calcul et de visualisation accessibles à tout chercheur prêt à se former à leur utilisation. Parmi les questions

ouvertes qui se posent dans le domaine figure la conception d'une méthodologie susceptible d'englober d'une manière cohérente le déploiement de mesures de calcul statistique (tests d'association, dispersion, productivité et asymétries co-occurrencelles) utilisées souvent en isolation d'une étude à l'autre.

L'usage des corpus en linguistique a connu un développement continu, qu'il s'agisse (i) de réunir les matériaux nécessaires à la description de langues peu documentées ; (ii) d'établir un corpus de référence pour le français, les dialectes et les langues de France ; (iii) de disposer de données de première main sur des questions spécifiques (acquisition et pathologies du langage, apprentissage d'une langue étrangère, variation phonologique, néologie, tests cognitifs, etc.) ; ou (iv) de construire scientifiquement les données nécessaires aux traitements informatiques et aux applications dans les industries de la langue. L'épreuve des hypothèses sur la structure des langues et leurs usages a été transformée par le recours à de grandes masses de données et l'exploitation de données orales. La fiabilité des transcriptions (conventions de notation, alignement sur le signal), la traçabilité par les métadonnées, l'assurance d'une disponibilité juridique et informatique, la conservation de la ressource initiale et de ses enrichissements ont mis en évidence la nécessité d'un *process* sous forme d'une chaîne de traitements contrôlés à chaque étape, robustes, libres de droit et pérennes. Les unités de recherche ont à leur disposition les moyens technologiques et l'expertise de l'IR\* Huma-Num (à travers le consortium Corpus, langues et interactions, CORLI), des plateformes comme COCOON (Collection de corpus oraux numériques) et les relais dans les MSH. Sont venus en renfort, dans le cadre des investissements d'avenir, l'Equipex ORTO-LANG (*Open Resources and Tools for Language*), et du côté du ministère de la Culture, les aides de la Délégation générale à la langue française et aux langues de France et le catalogage et l'hébergement assurés par la BnF.

Le recours aux corpus – qui constituent un élément fort d'identification des unités de recherche – est aujourd'hui intégré dans la

pratique des chercheurs. D'un côté, la compréhension des enjeux de la collecte et l'interprétation des productions langagières impliquent une collaboration avec l'anthropologie et les sciences sociales ; de l'autre, entre les activités de modélisation et de formalisation et la constitution et l'analyse des données, l'interface avec le TAL et l'IA, transgresse les frontières disciplinaires. Les chercheurs et chercheuses en LC sont toujours confrontés à trois défis majeurs : (i) la pérennité de la conservation des données dont la croissance (et donc les coûts de maintenance) est exponentielle ; (ii) la préservation de la capacité à développer une recherche publique dans un domaine où les applications commerciales conduisent à des investissements financiers considérables de la part de grandes entreprises ; et (iii) le développement des réseaux d'échange internationaux (comme par exemple, un engagement résolu dans les infrastructures européennes telles que Dariah et Clarin).

La LI est motivée par des questions plus linguistiques, et suit généralement les niveaux d'analyse classiques : phonologie, morphologie, syntaxe, sémantique, pragmatique, avec des questions spécifiques à chacun, par exemple produire des grammaires formelles à même de rendre possible l'analyse syntaxique automatique. Le TAL était historiquement peu distinguable de la LI et de ses découpages, se focalisant de plus sur des problématiques directement applicatives (la traduction automatique, les systèmes de dialogue humain-machine, l'extraction de connaissances à partir de textes) en partant du principe que la compréhension du sens était au cœur de la plupart des problèmes de TAL, et impliquait de résoudre aussi les problèmes posés par les niveaux intermédiaires.

Les travaux en sémantique distributionnelle ont connu une percée avec des représentations vectorielles capturant les similarités de sens, d'abord statique (un mot a une représentation unique), puis contextuelle (la représentation du mot est influencée par le contexte spécifique dans lequel il apparaît), avec plusieurs modèles très populaires apparus en 2017-2018 (BERT, GPT) qui ont rapidement fait

progresser tous les problèmes du TAL. Ces modèles reposent sur des approches prédictives : un modèle apprend automatiquement quel mot est le plus approprié dans un contexte donné, et construit une représentation numérique du mot qui permet de répondre à cette tâche, ce qui peut se combiner ensuite avec des entraînements spécifiques pour telle ou telle tâche. Le résultat est un ensemble de modèles complexes entraînés sur de grands corpus de textes et qui semblent avoir digéré un ensemble de contraintes linguistiques de façon implicite, sans représentations intermédiaires claires, et qui sont ensuite utilisés avec des ajustements spécifiques sur tous les problèmes du TAL.

Cette évolution a entraîné un schisme grandissant entre les préoccupations du TAL et celles de la LI, ainsi qu'un changement important de culture des communautés, la plupart des praticiens du TAL étant plutôt maintenant des experts en apprentissage automatique. L'accélération inédite du domaine vers plus de gigantisme dans les données et les calculs nécessaires a abouti depuis 2021-2022 à des modèles focalisés vers la génération de texte, qui semblent même pouvoir aborder les tâches les plus diverses de façon purement linguistiques, comme des oracles dont les prédictions ne dépendent que de la formulation des questions qu'on leur pose.

L'impact du TAL dans la société est donc maintenant quotidiennement visible et interroge sur de nombreux points qui dépassent la linguistique : mésusage de productions textuelles artificielle (plagiat, pollution informationnelle), captation des ressources et des moyens par une poignée d'acteurs privés travaillant de façon opaque et aux limites des cadres légaux existants (notamment par rapport au droit d'auteur). Mais les limites de ces modèles, qui préoccupent pour une grande part la communauté TAL, sont aussi des opportunités pour mobiliser les compétences des chercheurs en science du langage.

Les succès récents du TAL masquent en effet des problèmes liés aux données utilisées pour les entraîner, dont la représentativité des corpus collectés, avec des biais plus ou moins

gênants, dont le plus évident est la focalisation sur quelques langues bien dotées. Le manque de diversité linguistique des travaux en TAL est un danger supplémentaire pour la préservation des langues, mais limite aussi les perspectives de modélisation en TAL.

Un autre sujet dont les sciences du langage commencent à s'emparer est celui de l'analyse du fonctionnement des modèles automatiques, car leur opacité limite leur portée, en l'absence d'explication ou de garantie de leur comportement. Des méthodologies se rapprochent de pratiques en linguistique expérimentale, voire en neurolinguistique sont utilisées pour étudier les réelles capacités linguistiques des modèles, déterminer leurs limites et peut-être stimuler des évolutions. Il est notamment crucial de comprendre comment les facultés de langage semblent se développer chez l'humain en étant exposé à seulement une fraction de ce qui est nécessaire pour développer les modèles automatiques.

## 2. Sociolinguistique, linguistique anthropologique

La sociolinguistique est aujourd'hui une discipline au caractère multipolaire qui regroupe des recherches développées selon une grande diversité d'approches théoriques et méthodologiques. L'intérêt commun à tous les chercheurs et chercheuses de ce domaine porte sur les dimensions sociales des pratiques et des usages linguistiques à tous les niveaux des systèmes (prosodique, phonologique, syntaxique, discursif, etc.), qu'il s'agisse de langues orales ou signées. Par exemple, certaines recherches plus généralement consacrées à l'étude des systèmes sonores, approfondissent des aspects spécifiques pertinents pour la sociolinguistique tels que la sociophonétique et la sociophonologie ou l'étude de la voix genrée.

Dans le domaine de la sociolinguistique, les données observables étant recueillies souvent sur le terrain, le travail d'enquête est central, sans pour autant négliger les productions écrites, dans le cadre d'une sociolinguistique des textes allant, par exemple, des biographies

migratoires aux écritures exposées dans les contextes urbains, ou encore aux productions des internautes. Les documents écrits sont aussi la source fondamentale de la sociolinguistique historique – approche encore en évolution en France – qui prend en compte la composante sociale dans l'étude diachronique des usages attestés dans des textes proches de l'oralité (par exemple, textes anciens écrits par des personnes peu lettrées, documents comptables, autobiographies, témoignages de procès).

La constitution de corpus surtout oraux, d'ampleur différente, constitue un enjeu important pour les sociolinguistes et notamment pour ceux qui se consacrent au milieu urbain, observatoire privilégié de la diversité et du contact aux niveaux socioculturel et linguistique. Dans cette direction, il faut mentionner, la mise en place de grands corpus variationnistes de sociolinguistique urbaine : Enquêtes sociolinguistiques à Orléans (ESLO), corpus prototypique hébergé sur Huma-Num ; *Multicultural Paris French* (MPF), relatif aux « parlers jeune » de la région parisienne ; bases de données sonores qui rendent compte de la situation sociolinguistique de Toulouse et de Marseille ; etc. D'autres corpus sont enregistrés dans des situations plus variées, notamment professionnelles.

Dans le cadre de la sociolinguistique variationniste, l'approche quantitative est prise en compte en fonction de la modélisation formelle et computationnelle des théories sociolinguistiques ou en fonction des recherches en sociolinguistique développementale en direction aussi d'un croisement avec les sciences cognitives. Dans le cadre de la sociolinguistique critique, l'approche qualitative est mise en avant, et le focus porte sur le rôle de la langue dans les rapports de domination et la reproduction sociale, et donc un croisement avec la sociologie et les études politiques.

### 3. Psycholinguistique, neurolinguistique, acquisition

La neurolinguistique concerne la compréhension de la relation entre cerveau et langage, ainsi que de l'impact du dysfonctionnement

cérébral sur le comportement langagier. Le maître-mot de son champ d'action est « réseaux », étudiés *via* le prisme de modèles linguistiques, psycholinguistiques, computationnels, neuropsychologiques et neurophysiologiques. Les recherches en neurolinguistique sont transversales, multi- et interdisciplinaires, et concernent la production, la perception et la compréhension de la parole et du langage (chez l'adulte), tout comme son acquisition (chez l'enfant dans le développement du langage, comme chez l'adulte dans l'apprentissage des langues secondes), sa perte et sa remédiation/réappropriation (chez le patient). La psycholinguistique, quant à elle, vise à décrire et à analyser les mécanismes cognitifs qui permettent de produire et comprendre le langage au sens large, et les représentations mentales et cérébrales du langage chez des individus présentant ou non une pathologie. Le maître-mot de son champ d'action est « processus ».

Les recherches en neurolinguistique et psycholinguistique recourent à des méthodes d'investigation expérimentale multiples et variées : comportementales et psychophysiques (par exemple, chronométrie mentale, oculométrie, enregistrements articulatoires et acoustiques), électrophysiologiques (EEG, SEEG), neuroimagerie (IRMf, TMS, NIRS, TEP). En fonction de la recherche, les protocoles utilisés privilieront une bonne résolution spatiale (c'est-à-dire la précision de la localisation des événements observés dans le cerveau, telles que les techniques de neuroimagerie), une bonne résolution temporelle (c'est-à-dire la précision du moment où se passe l'événement, telle que l'EEG), ou une méthode/analyse permettant de combiner les deux. La modélisation computationnelle se basant sur de grandes quantités de données est en plein essor dans le domaine.

Les travaux réalisés en psycholinguistique et neurolinguistique portent principalement sur le traitement du langage parlé, tant en perception qu'en production, depuis les traitements auditifs, phonétiques et phonologiques jusqu'à la syntaxe, la sémantique et la pragmatique en passant par la morphologie et l'accès au lexique. Les mécanismes de traitement du

langage parlé sont étudiés en interaction avec les caractéristiques du locuteur (par exemple, bilingue, multilingue, monolingue). Par ailleurs, les recherches interlangues connaissent un développement important, cherchant à la fois à déterminer des processus universaux langagiers communs à toutes les langues et traités de manière indépendante de la langue du locuteur, et des spécificités de langue qui auraient une dynamique cérébrale particulière. L'ensemble de ces travaux vise notamment à étudier l'impact des propriétés spécifiques des systèmes linguistiques sur les mécanismes de traitement du langage. Les recherches sur le traitement du langage écrit portent sur la lecture, mais aussi sur le processus d'écriture en tant que geste moteur spécifique langagier, l'influence de l'orthographe sur la perception phonétique ou l'accès au lexique, sur les processus de médiation phonologique et sur l'analyse morphologique. Quelques travaux traitent plus particulièrement du lien production-perception tant au niveau du langage parlé que du langage écrit.

« L'approche pathologique » est intimement liée aux recherches en neuro- et psycholinguistique, qui prennent leurs fondements dans les descriptions anatomocliniques de patients souffrant de lésions impactant parole et langage. Les travaux impliquant différentes populations cliniques (dyslexie, dysphasie, individus malentendants avec ou sans implant cochléaire, affections neurodégénératives [telles que maladie de Parkinson ou démence de type Alzheimer], lésions cérébrales, autisme, schizophrénie) et l'impact des différentes pathologies sur les mécanismes de traitement du langage et de la parole (syntaxe, sémantique, pragmatique, phonétique, phonologie, etc.) représentent une part conséquente des investigations et du développement interdisciplinaires au sein des sciences du langage. La remédiation, la réappropriation, les neurochirurgies fonctionnelles et les stratégies compensatoires à l'oral et à l'écrit sont autant d'exemples d'études interventionnelles dont le fort impact sociétal/clinique est en plein développement dans la section 34. Sur ce point, les travaux entrant dans le cadre d'une recherche appliquée permettent par exemple

la création d'outils diagnostiques, de bilan ou de remédiation/rééducation auprès de populations, enfants ou adultes, souffrant de troubles linguistiques.

Une part importante des recherches en neuro- et psycholinguistique est dédiée à l'acquisition du langage chez le nourrisson et le jeune enfant ayant un développement typique ou atypique. Cette thématique de recherche bénéficie d'un réel essor, comme en témoignent par exemple les récents coloriages de postes au concours CRCN CNRS (cf. tableau 3), le développement d'équipes de recherche dans les laboratoires et les nombreux financements institutionnels et nationaux dédiés. L'acquisition du langage est également étudiée chez l'adulte, dans le cas notamment de l'apprenant d'une langue seconde. Le précédent rapport de conjoncture relevait des interactions de plus en plus nombreuses de ce champ thématique avec d'autres champs des sciences cognitives et de la linguistique. Outre les interactions déjà anciennes avec la phonologie et la phonétique, il était rapporté une ouverture en direction de la sémantique formelle, de la modélisation informatique et de la primatologie (comparaisons des capacités d'apprentissage du nourrisson ou du jeune enfant avec d'autres espèces de primates non humains). Il apparaît que ces épanouissements thématiques sont toujours d'actualité.

Les recherches en psycholinguistique et neurolinguistique impliquent des laboratoires rattachés de manière principale à la section 34, mais également des laboratoires de rattachement secondaire (principalement section 26, parfois 7 ou 35) ayant dans ce cas développé un axe de recherche identifié « langage ». Ces recherches résultent ainsi d'interactions entre sciences du langage et sciences cognitives, biologiques, informatiques et théoriques.

#### **4. Histoire des théories linguistiques, épistémologie**

L'étude réflexive des conceptualisations de la linguistique, appréhendées dans leur dimension historique, se réalise de trois façons différentes dans le panorama de la recherche. Le

domaine de l'histoire des théories linguistiques (i) participe à la mise en perspective de travaux contemporains qui sont saisis comme le résultat d'un progrès dans la discipline / le domaine d'application, à l'intérieur d'unités de recherche qui se consacrent premièrement à d'autres domaines ; (ii) restitue les travaux de phonétique, de lexicographie, d'histoire des systèmes d'écriture ou de grammatisation et production de grammaire d'écoles qui se sont développées en dehors de la tradition occidentale (écoles sanskrites, arabes, chinoises, etc.), au sein d'équipes qui se consacrent à l'analyse des langues qui ont forgé ou emprunté ces écritures ; (iii) s'attache à la dimension proprement historique des théories.

Les recherches portent sur l'histoire des conceptions du langage et des langues, ainsi que sur la réflexion épistémologique et socio-ologique qui accompagne les développements des études en sciences du langage, sur un arc chronologique qui va de l'Antiquité à l'époque contemporaine et rassemble des spécialistes d'aires linguistiques et de domaines conceptuels variés.

Parmi les thèmes porteurs, on peut citer celui des « grammaires étendues », consacré à l'extension des modèles grammaticaux sanskrit, grec, arabe, mais aussi, plus récemment, aux interactions durables et aux emprunts entre traditions grammaticales à travers les siècles (notamment interactions entre grammaire grecque, syriaque et arabe entre l'Antiquité tardive et l'époque abbaside). La glossographie latine médiévale a donné lieu à de nombreux projets interdisciplinaires (numérisation et analyse computationnelle de corpora textuels, création de bases de données en libre accès, conservation du patrimoine linguistique et manuscrit) et collaborations internationales. Des éditions de textes grammaticaux issus de traditions différentes sont produites régulièrement (corpus d'Apollonius Dyscole, Priscien et ses glossateurs, grammairiens français des XVII<sup>e</sup> et XVIII<sup>e</sup> siècles).

La recherche française joue un rôle de premier plan dans les études des Grammaires missionnaires (XVI<sup>e</sup>-XIX<sup>e</sup> siècles, histoire de la description des langues d'Asie, des Amériques,

d'Afrique, d'Océanie), ainsi que dans le domaine de l'ethnolinguistique. La linguistique moderne et contemporaine fait aussi l'objet d'études dans les domaines suivants : corpus saussurien et théories saussuriennes du signe, théories du langage dans les sciences cognitives, histoire de l'automatisation des sciences du langage. Dans le domaine de l'épistémologie, des études sont menées sur le cognitivisme, sur la biosémiose, sur la grammatisation dans ses implications dans la modélisation grammaticale et description des langues dans les différentes approches contemporaines.

## C. Variations et diversités

### 1. Évolution du langage

L'évolution du langage dans l'espèce *Homo sapiens* est une question par nature profondément interdisciplinaire, à la frontière de la biologie, de la paléoanthropologie, des sciences du langage et des neurosciences cognitives. On peut distinguer deux grands types d'approches parmi les théories actuelles sur l'évolution du langage. Il y a, d'une part, des approches fondées sur la simulation informatique de processus comme la création de conventions lexicales, l'étude des conditions nécessaires à l'évolution de la coopération (souvent jugée comme indispensable à l'apparition du langage) ou les limites de la communication holistique qui rendent indispensable la double articulation pour un système de communication productif, comme l'est le langage. Ces approches sont souvent basées sur la théorie (évolutive) des jeux. Il y a, d'autre part, des approches de nature plus théorique et interdisciplinaire, qui cherchent à identifier les spécificités structurelles du langage utilisé dans la communication humaine et ce qui lui est propre par rapport aux systèmes de communication animale (psychologie comparée). Ces spécificités liées aux capacités cognitives et culturelles humaines permettent de dégager des scénarios pour l'évolution du langage. Ce deuxième courant se base sur les neurosciences cognitives, la psychologie comparée,

la (paléo-)anthropologie, la simulation informatique et les sciences du langage.

## 2. Linguistique historique

L'étude du changement des langues à travers le temps et des parentés entre les langues est l'une des branches les plus anciennes de la linguistique. Les méthodes de reconstruction et de classification des langues ainsi que les hypothèses sur les différentes familles de langues, si elles ont progressé au fil des découvertes et des croisements avec la linguistique générale, la géographie linguistique et la sociolinguistique, sont restées relativement stables depuis le xix<sup>e</sup> siècle, et font ainsi de la linguistique historique et comparative un domaine remarquablement cumulatif et unifié au sein de la linguistique. Les recherches comparatives se situent souvent à l'interface avec la linguistique de la diversité et la typologie des langues et sont souvent menées par des linguistes de terrain sur des familles de langues diverses.

La linguistique historique et comparative connaît ces dernières années des avancées nouvelles grâce à une grande ouverture à l'interdisciplinarité. Les travaux sur les classifications des langues ont progressé et ont connu un renouveau grâce à l'apport des méthodes phylogénétiques computationnelles utilisées en biologie. En outre, les collaborations entre linguistes, archéologues et généticiens ont permis des avancées importantes sur l'histoire des populations humaines, de leurs mouvements, et de leurs interactions. La linguistique historique et comparative bénéficie également du développement des humanités numériques, de la numérisation de corpus écrits de nombreuses langues anciennes et de la mise à disposition de bases de données sur les langues qui facilitent le test et la réplicabilité des hypothèses, et permettent d'appliquer des méthodes statistiques et computationnelles. La création de bases de données comparatives et étymologiques en ligne, augmentées et améliorées de façon continue permettent de dépasser les limites imposées par les dictionnaires traditionnels. D'autre part, les études récentes

sur les changements grammaticaux (en phonologie, syntaxe et sémantique) ont bénéficié d'un nouvel élan grâce à de nouvelles connaissances théoriques : on peut citer, à titre d'exemple, les recherches sur la grammaticalisation, ainsi que les approches génératives qui soulignent le rôle des apprenants de première langue comme agents du changement linguistique (en particulier pour les processus dits endogènes). D'autres avancées ont lieu grâce aux interactions mutuellement bénéfiques avec les autres domaines de la linguistique comme entre autres la sociolinguistique, les études sur les créoles, la linguistique aréale et l'étude du contact des langues.

Dans les prochaines années, on peut prévoir que la linguistique historique et comparative va continuer son ouverture vers l'utilisation de modèles statistiques et computationnels, ainsi que l'informatisation de ses méthodes de comparaison et de reconstruction, tout en continuant à jouer un rôle majeur dans les travaux interdisciplinaires sur l'histoire des populations.

## 3. Typologie, diversité, variation linguistique

La typologie linguistique étudie la diversité de traits linguistiques attestés à travers les langues du monde ainsi que leur variation et leur évolution dans le temps. Elle porte sur tous les niveaux : phonétique, phonologique, morphologique, syntaxique, sémantique, pragmatique, discours, contact de langues, etc. La typologie s'intéresse aussi à la description de langues peu étudiées, dans des perspectives variées : fonctionnaliste, linguistique théorique, etc. La description typologique des langues est aussi liée à un courant en pleine expansion, celui de la documentation linguistique qui met l'accent sur le recueil et l'archivage pérenne de données linguistiques riches, permettant de préserver un témoignage des langues menacées et d'assurer leur disponibilité pour d'éventuelles descriptions et recherches futures. Une autre tendance récente est l'adoption d'approches typologiquement informées par la philologie et la linguistique historique.

Si les bases de données qui recensent les propriétés d'un nombre de langues du monde relativement grand, comme le *World Atlas of Language Structures*, ont marqué la typologie dans les années 2000, leurs limites ont mené au développement de nouvelles approches, fondées sur des bases de données plus réduites mais mieux maîtrisées. Sur le plan méthodologique, la typologie des langues est confrontée aux problèmes inhérents du recensement de phénomènes linguistiques à travers des grammaires rédigées par divers auteurs et à différents moments de l'histoire de la discipline. En réponse à ces difficultés, les nouvelles grammaires tendent à être associées à des corpus de données orales et écrites. De nombreux travaux sont fondés sur des questionnaires typologiques associés à des stimuli conçus pour l'étude de phénomènes spécifiques tout en étant culturellement adaptés aux différentes populations à travers le monde (par exemple, stimuli vidéo). On note d'autre part un renouvellement dans le domaine de la typologie quantitative qui s'appuie de plus en plus sur des modèles statistiques sophistiqués et une intensification des collaborations avec l'anthropologie moléculaire, les neurosciences ou la géographie.

#### 4. Dialectologie, contact de langues

L'interaction entre différentes langues dans les situations multilingues peut affecter les systèmes linguistiques et/ou leurs usages, à des niveaux multiples : choix de langue, alternance codique, transferts, emprunts, ou encore émergence de nouvelles variétés langagières (créoles, pidgins, variétés de contact). Dans ce contexte, un large éventail d'approches typologiques ou théoriques s'intéresse aux dynamiques de contact et aux langues créoles, et la recherche s'étend aussi aux disciplines connexes, telles que les neurosciences, la sociolinguistique, la psycholinguistique, la psychologie comparée, l'ethnologie.

La diversité et la variation linguistique (phonologique, syntaxique, morphologique, etc.) se retrouvent à l'intérieur des langues individuelles. La dialectologie, qui permet de cibler

de manière précise et systématique ces points de microvariation, contribue de manière fondamentale à la typologie et à la compréhension des mécanismes linguistiques. La production d'atlas linguistiques cartographiés connaît par ailleurs un développement important, grâce à des collaborations interdisciplinaires entre dialectologues et chercheurs en informatique et en géomatique.

#### 5. Langues signées

Il est aujourd'hui acquis que les langues des signes, qui utilisent la modalité visuo-gestuelle et sont spontanément créées par les personnes sourdes, sont des langues naturelles au même titre que les langues vocales. La linguistique des langues des signes est un domaine de recherche établi, disposant de ses propres revues et conférences, et nombre de colloques de linguistique intègrent une ou plusieurs contributions sur ces langues. Deux grands types d'approches se distinguent actuellement en France. Une question centrale leur est toutefois commune, qui est de savoir quelle est la part de propriétés linguistiques universelles, c'est-à-dire partagées par les langues vocales et les langues des signes, et celle de propriétés éventuellement spécifiques à chaque ensemble de langues, notamment en raison de la différence de modalité, audio-phonatoire d'un côté, visuo-gestuelle de l'autre.

Le premier type d'approche visant à établir que les langues des signes sont des langues naturelles, a prioritairement cherché, dans ces langues, la contrepartie des propriétés fondamentales déjà décrites pour les langues vocales. Cette phase a été initiée par des travaux sur la langue des signes américaine, en particulier ceux de W. Stokoe (1960), suivis, une décennie plus tard, par ceux d'un groupe de chercheurs inscrits dans le cadre du paradigme générativiste initié par Noam Chomsky. Une deuxième approche initiée par le linguiste Christian Cuxac et connue sous le nom de « modèle sémiologique », s'attache à mettre en relation les fonctions remplies par les discours en langue des signes (raconter une histoire,

expliquer, argumenter, etc.) et leurs caractéristiques structurales. Un point traditionnel de discorde entre les deux paradigmes concerne la centralité de la notion d'iconicité, mais les travaux plus récents indiquent que cette question doit être explorée (et admet des réponses partiellement différentes) en faisant des distinctions entre le niveau lexical et les différents niveaux de la phrase (par exemple, entre structures avec et sans classificateurs).

Au-delà des recherches sur la structure, l'acquisition et la dimension sociolinguistique des langues des signes, un nouvel axe de recherche s'est ajouté au paysage français depuis quelques années, celui qui concerne la psycholinguistique des langues des signes.

## V. Verrous actuels et défis prospectifs

À la suite d'une journée (en septembre 2023) réunissant les directeurs et directrices d'unités rattachées de manière principale ou secondaire à la section 34, invités à engager avec nous un dialogue collégial d'état des lieux, nous avons identifié conjointement les verrous et les obstacles, mais également les défis et les ouvertures auxquels notre communauté scientifique fait actuellement face.

### A. Verrous actuels

#### 1. Le contexte préoccupant de la linguistique théorique

La linguistique fondamentale fait face à deux problèmes majeurs : (i) le lent mais tangible déclin des recherches théoriques ; et (ii) le passage de la linguistique théorique de discipline cœur de domaine à discipline assujettie à d'autres sciences du langage, en particulier celles utilisant des méthodes expérimentales et/ou corpus. Ces deux constats interviennent au moment de l'entrée en force

des LLM (*Large Language Models*), qui demandent, pour être compris et maîtrisés, que l'on poursuive les efforts théoriques pour la compréhension de la faculté du langage humain. Parmi les sciences du langage, il semble que les autres sous-disciplines ne sont pas sujettes à des difficultés d'une même ampleur.

Cette désertion de la théorie peut s'expliquer par deux constats faits par les chercheurs et chercheuses de la section : (i) qu'il est de plus en plus difficile d'obtenir des financements pour des recherches théoriques (par exemple, les projets purement théoriques sont très rarement financés par l'ANR ; et il est même difficile de trouver des comités d'experts pour évaluer les projets) ; et (ii) qu'il est difficile, sinon impossible, de trouver des étudiants formés aux études théoriques et de les financer.

La lente désertion des approches théoriques n'est pas unique et propre à la linguistique. Le constat d'une diminution d'intérêt pour la recherche fondamentale est partagé par plusieurs sections du CNRS. Spécifiquement pour l'étude du langage, les études théoriques servent à identifier les règles essentielles de la faculté propre à l'être humain et à les modéliser. L'identification de ces règles est issue d'un effort conjoint des approches théoriques avec les apports de la psycholinguistique, la typologie contrastive et la linguistique de corpus. Mais sans la théorie, il n'y a pas de modélisation ni d'abstraction.

La solution à ce premier problème est multifactorielle : d'une part, il faudrait augmenter le nombre de théoriciens auprès des comités d'évaluations de projets. Il faudrait aussi continuer à développer les relations entre recherche et formation initiale, par exemple en développant les cours théoriques dès le début de la formation en linguistique dans les universités, ce qui se fait déjà mais pas partout. La recherche dans les laboratoires peine parfois à faire le lien avec la formation. D'autant plus que les universités recrutent depuis plusieurs années essentiellement en didactique et en français langue étrangère, sans que la recherche en sciences du langage sur ces thématiques voit le même engouement. De plus, les écoles doctorales pourraient aussi jouer un rôle en

orientant des bourses de thèse à destination précisément des recherches fondamentales non appliquées.

Par ailleurs, la linguistique théorique devient trop souvent une discipline ancillaire, alors qu'elle est centrale quant à la compréhension et au développement de l'IA et des enjeux liés aux LLM. La section 34 peine à s'approprier le TAL à sa juste mesure, contrairement à la section 7 et à la CID 55. Cette situation difficile est aussi reflétée dans l'importance relative que recouvre l'activité de valorisation dans l'évaluation des personnels, alors que c'est une partie significative de l'activité des TAListes. Une situation comparable peut être évoquée pour la psycholinguistique : son appartenance thématique à la fois dans les sections 34 et 26 est à la fois un atout, mais engendre également des déséquilibres et un manque de clarté quant à la visibilité de la discipline. Les relations entre sections et commissions interdisciplinaires ont été identifiées comme un point de vigilance.

## 2. La relation entre recherche et formation

Un autre verrou concerne les relations entre laboratoires (unités mixtes) et le monde universitaire. De plus en plus de laboratoires de recherche en section 34 peinent à faire le lien avec la formation. De moins en moins d'étudiantes et étudiants se tournent vers les disciplines liées à la section 34, possiblement par crainte de ne pas trouver de postes ou d'emploi ; qui plus est, le nombre de postes ouverts au concours CNRS en section 34 est en constante baisse depuis 20 ans. Nous regrettons une concertation encore insuffisante entre les tutelles des UMR en ce qui concerne la conjoncture scientifique et les besoins en termes de profils (universitaires) et de coloriages (CNRS) de postes, ce qui pourrait grandement faciliter le recrutement de chercheurs et chercheuses à même de relever les défis actuels en sciences du langage.

## 3. Des obstacles communs au monde la recherche

Pour finir, rappelons que le Conseil scientifique du CNRS a synthétisé de manière

concrète le sentiment auparavant latent d'une « exaspération croissante à l'encontre d'un alourdissement régulier du cadre administratif » au travers d'un *Livre blanc préliminaire sur les entraves à la recherche – Focus entraves administratives* (mai 2023). De fait, tous les chercheurs et chercheuses sont amenés à consacrer une part de leur temps de travail à des tâches administratives lourdes et souvent chronophages. Le rapport de la Commission d'études spécialisées portant sur la recherche, adopté par le CNESER le 4 juillet 2023 et intitulé *Transformation de la gestion du temps de travail et des activités dans la recherche publique et financement par appels à projets*, confirme que l'opinion dominante est que l'ampleur et la complexité de ces tâches administratives s'accroissent, impactant de manière négative l'activité de recherche.

## B. Défis prospectifs et thématiques émergentes

Un constat s'impose pour les sciences du langage : il est nécessaire de cultiver l'interdisciplinarité, tout en préservant les fondamentaux disciplinaires. Un objectif est de trouver les moyens d'être plus visible en tant que discipline, notamment au sein des SHS, en adoptant, par exemple, une politique de science par et pour la société.

Les disciplines de la section 34 doivent continuer à trouver les moyens de s'ouvrir à d'autres disciplines. Mettre en place des collaborations entre plusieurs domaines permettra d'ancrer notre section et de l'ouvrir à d'autres approches et fondements. La section doit continuer dans sa politique de collecte de données et poursuivre dans une démarche expérimentale, notamment pour les langues peu dotées ou en danger. Par exemple les pratiques issues du TAL, qui est depuis son essor en France rattaché à la linguistique, pourraient aider à la conservation et la valorisation de langues spécifiques. En d'autres termes, en développant une approche interdisciplinaire, certaines disciplines pourraient ne plus être menacées et

bénéficier d'un renouveau. De fait, l'un des secteurs au sein desquels la linguistique est attendue est bien évidemment celui du TAL, mais l'apport de la recherche en TAL n'est pas suffisamment mis en valeur au sein de la section. Un effort doit être fait en ce sens.

On note aussi que la linguistique a une position hybride au sein des SHS. Alors qu'elle appartient à cet institut, elle partage des pratiques de publications avec les sciences dures, ce qui la rend parfois peu ou mal reconnaissable au sein de l'institut qui est le sien. D'autre part, elle est ignorée par les informaticiens qui la considèrent comme ancillaire. Cette place hybride doit et peut être transformée en une force avec une communication plus claire, insistante et une mise en valeur avec des actions diverses autour de la discipline. Ceci est d'autant plus important que les pratiques de publication dans des cadres interdisciplinaires sont difficiles, les supports adaptés n'étant pas nombreux.

Enfin, deux notions sont incontournables dans les défis à venir pour les sciences du langage : la mise en place de plateformes de récolte et de partage de données à grande échelle, et la valorisation environnementale et sociétale de nos recherches, afin de s'inscrire dans une démarche forte de science ouverte et partagée, bénéficiant de la mise en commun des moyens.

## 1. Une politique favorable du CNRS

Certains éléments de politique scientifique du CNRS sont en cours et sont importants pour la valorisation de notre section. On pourrait citer, par exemple : le dialogue entre CNRS Sciences humaines & sociales et CNRS Biologie autour des thématiques scientifiques émergentes ; la mise en place d'une feuille de route TAL entre CNRS Sciences humaines & sociales et CNRS Sciences informatiques ; le Contrat objectifs performance (COP) ; un traitement spécifique de la langue française et des langues de France, y compris la langue des signes ; le renouvellement probable du COP Défi inégalités éducatives, qui est un observatoire des

inégalités éducatives et qui concerne CNRS Sciences humaines & sociales, CNRS Biologie et CNRS Sciences informatiques. Enfin, des outils et dispositifs mis en place actuellement tels que les IRL de Chicago et de São Paulo, qui concernent les sections 34 et 35, ou encore le SOSI (Suivi ouvert des sociétés et de leurs interactions) de CNRS Sciences humaines & sociales sont autant de possibilités que chercheurs et chercheuses de la section 34 doivent s'approprier.

## 2. Thématiques émergentes

Toutes les recherches en lien avec le TAL et l'IA sont identifiées comme des atouts importants pour notre section (pour plus de détails, cf. partie IV.B.1). À la question posée aux directeurs et directrices des UMR rattachées à la section 34, au sujet de leurs compréhensions des thématiques émergentes qui, selon elles/eux, devront être développées dans les dix prochaines années, ont été évoquées :

- les études aréales (qui représentent une priorité de CNRS Sciences humaines & sociales) ;
- les données en SHS ;
- l'anthropologie linguistique ;
- la psycholinguistique ;
- les langues peu dotées, y compris les langues des signes ;
- les langues régionales de France en lien avec le TAL ;
- la prise en charge de la migration des populations dans le cadre des sciences du langage ;
- la linguistique expérimentale et les larges cohortes.

Ces thématiques sont autant d'exemples d'enjeux actuels, ancrés dans la société, qui attestent d'une section en phase avec les besoins multidisciplinaires de ses thématiques.

## VI. Conclusion

La section 34 est une section dynamique. L'activité des chercheuses et des chercheurs – y compris ingénieries/ingénieurs de recherche – qui la composent est excellente et reconnue. C'est une section en constante évolution, adaptation, renouvellement en termes de méthodes et concepts théoriques nouveaux, au service de la compréhension d'un objet de recherche fondamentale de notre société : le langage.

L'évolution des structures participant à l'organisation de la recherche de certaines sous-disciplines, a été bénéfique : on peut citer par exemple les succès des Labex EFL (Paris) et ASLAN (Lyon), de l'institut Convergences ILCB (Aix-Marseille), et le remodelage du paysage scientifique par le biais des grandes universités de recherche. Ces niveaux

organisationnels participent à fédérer la communauté des sciences du langage, tout comme les initiatives tels que les IRN et le réseau thématique prioritaire Éducation.

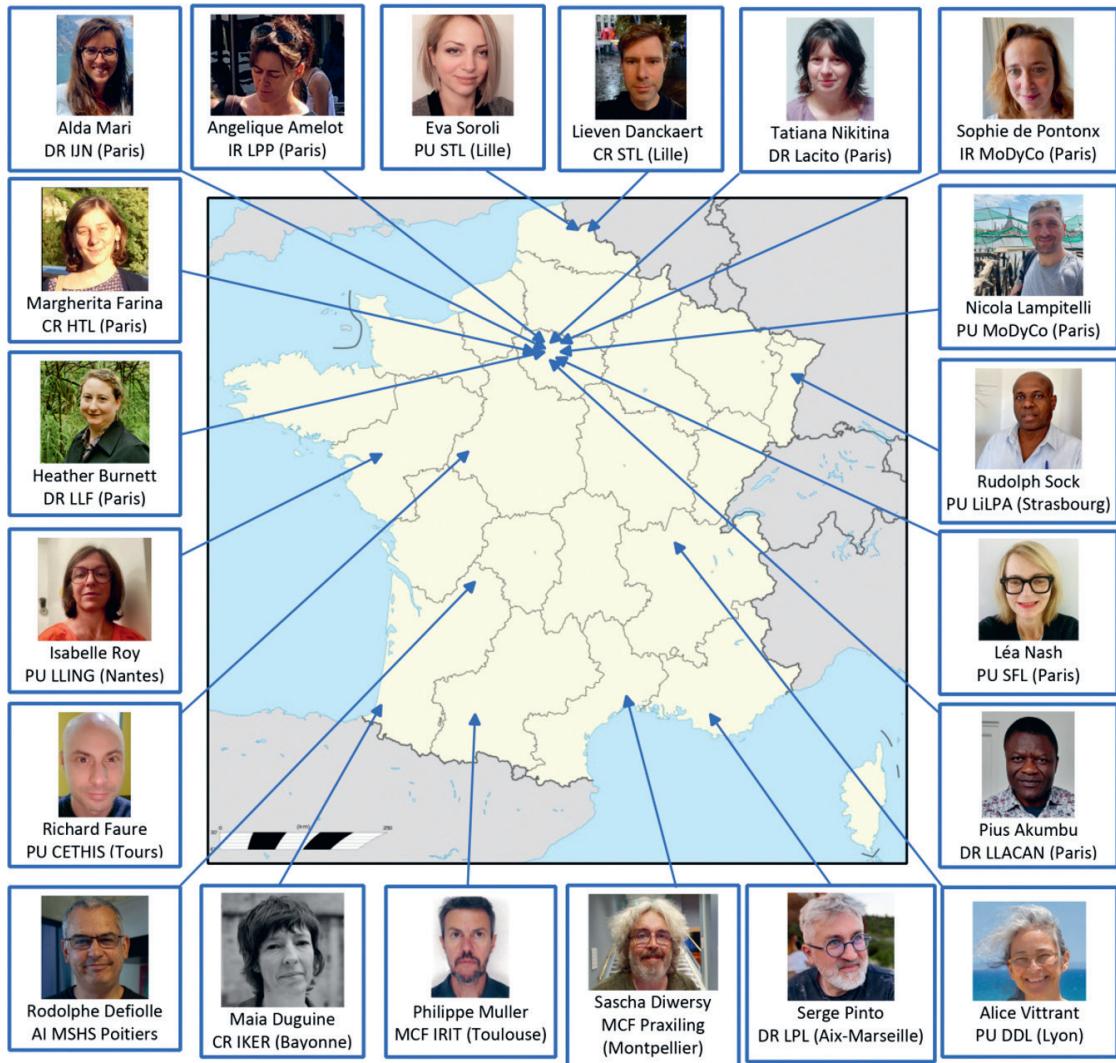
Il faut néanmoins rappeler l'impact des confinements de 2020 et 2021 lors de la crise planétaire liée à la Covid-19 qui a considérablement freiné toute initiative, autant dans la recherche en sciences du langage, que dans la recherche en général, tant d'un point de vue concret (moins d'activité) que conceptuel (réflexion sur le sens même et la finalité de la recherche). Cet impact reste présent lors de la rédaction de ce rapport.

Pour conclure, nous voulons remercier nos prédécesseurs, les membres des mandatures passées, qui ont produit des rapports de conjoncture qui ont été pour nous source d'inspiration et sur lesquels nous nous sommes fondés pour la rédaction de ce document.

Comité national de la recherche scientifique

## ANNEXE 1

### Trombinoscope de la section 34, mandat 2021-2025



---

## ANNEXE 2

### Laboratoires rattachés à la section 34 de manière principale (décembre 2023)

---

Code unité	Intitulé	Sigle	Ville	Région	Chercheurs et chercheuses CNRS rattachés à la section 34	Chercheurs et chercheuses CNRS rattachés à d'autres sections
UMR 5191	Interactions, Corpus, Apprentissages, Représentations	ICAR	Lyon	Sud-Est	3	1 (section 26)
UMR 5263	Cognition, Langues, Langage, Ergonomie	CLLE	Toulouse	Sud-Ouest	6	1 (section 26)
UMR 5267	PRAXILING	PRAXILING	Montpellier	Sud-Ouest	0	0
UMR 5478	Centre de Recherche sur la langue et les textes basques	IKER	Bayonne	Sud-Ouest	7	0
UMR 5596	Dynamique Du Langage	DDL	Lyon	Sud-Est	8	2 (sections 26 et 34) 1 (section 26)
UMR 6310	Laboratoire de Linguistique de Nantes	LLING	Nantes	Nord-Ouest	6	0
UMR 7018	Laboratoire de Phonétique et Phonologie	LPP	Paris	Île-de-France	7	0
UMR 7023	Structures Formelles du Langage	SFL	Saint-Denis	Île-de-France	10	0
UMR 7107	Laboratoire de Langues & Civilisations à Tradition Orale	LACITO	Villejuif	Île-de-France	5	1 (section 38)
UMR 7110	Laboratoire de Linguistique Formelle	LLF	Paris	Île-de-France	8	1 (section 26)
UMR 7114	Modèles, Dynamiques, Corpus	MoDyCo	Nanterre	Île-de-France	1	0

## Comité national de la recherche scientifique

Code unité	Intitulé	Sigle	Ville	Région	Chercheurs et chercheuses CNRS rattachés à la section 34	Chercheurs et chercheuses CNRS rattachés à d'autres sections
UMR 7118	Analyse et Traitement Informatique de la Langue Française	ATILF	Nancy	Grand-Est	3	0
UMR 7270	Laboratoire Ligérien de Linguistique	LLL	Orléans	Nord-Ouest	0	0
UMR 7309	Laboratoire Parole et Langage	LPL	Aix-en-Provence	Sud-Est	17	3 (section 26) 1 (section 7)
UMR 7320	Bases, Corpus, Langage	BCL	Nice	Sud-Est	7	0
UMR 7597	Histoire des Théories Linguistiques	HTL	Paris	Île-de-France	6	1 (section 35)
UMR 8094	Langues, Textes, Traitements informatiques, Cognition	LATTICE	Montrouge	Île-de-France	8	0
UMR 8135	Langage, Langues et Cultures d'Afrique Noire	LLACAN	Villejuif	Île-de-France	11	0
UMR 8202	Structure et Dynamique des Langues	SeDyL	Villejuif	Île-de-France	5	0
UMR 8563	Centre de Recherches Linguistiques sur l'Asie Orientale	CRLAO	Paris	Île-de-France	9	0
				<b>Total</b>	<b>127</b>	<b>12</b>

---

## ANNEXE 3

### Laboratoires rattachés à la section 34 de manière secondaire (décembre 2023)

---

<b>Code unité</b>	<b>Intitulé</b>	<b>Sigle</b>	<b>Ville</b>	<b>Région</b>	<b>Section de rattachement principale</b>	<b>Chercheurs et chercheuses CNRS rattachés à la section 34</b>
UMR 5105	Laboratoire de Psychologie et de Neurocognition	LPNC	Saint-Martin-d'Hères	Sud-Est	26	1
UMR 5216	Grenoble Images Parole Signal Automatique	GIPSA-lab	Saint-Martin-d'Hères	Sud-Est	7	6
UMR 5229	Institut des Sciences Cognitives Marc Jeannerod	ISC-MJ	Bron	Sud-Est	26	2
UMR 6024	Laboratoire de Psychologie Sociale et Cognitive	LAPSCO	Clermont-Ferrand	Sud-Est	26	0
UMR 7170	Institut de Recherche Interdisciplinaire en Sciences Sociales	IRISSO	Paris	Île-de-France	40	0
UMR 7295	Centre de Recherches sur la Cognition et l'Apprentissage	CeRCA	Poitiers	Nord-Ouest	26	1
UMR 7503	Laboratoire Lorrain de Recherche en Informatique et ses Applications	LORIA	Vandoeuvre-lès-Nancy	Grand-Est	6	3
UMR 8041	Centre de recherche sur le monde iranien : Langues, cultures et sociétés de l'Antiquité à nos jours	CeRMI	Ivry-sur-Seine	Île-de-France	32	1
UMR 8129	Institut Jean-Nicod	IJN	Paris	Île-de-France	35	8

## Comité national de la recherche scientifique

Code unité	Intitulé	Sigle	Ville	Région	Section de rattachement principale	Chercheurs et chercheuses CNRS rattachés à la section 34
UMR 8132	Institut des Textes et Manuscrits modernes	ITEM	Paris	Île-de-France	35	1
UMR 8163	Savoirs, Textes, Langage	STL	Villeneuve-d'Ascq	Nord-Ouest	35	4
UMR 8546	Archéologie et philologie d'Orient et d'Occident	AOROC	Paris	Île-de-France	31	0
UMR 8554	Laboratoire de sciences cognitives et psycholinguistique	LSCP	Paris	Île-de-France	26	3
UMR 8590	Institut d'Histoire et de Philosophie des Sciences et des Techniques	IHPST	Paris	Île-de-France	35	0
UMR 9015	Laboratoire Interdisciplinaire des Sciences du Numérique	LISN	Saint-Aubin	Île-de-France	7	3
UMR 9193	Sciences Cognitives et Sciences Affectives	SCALab	Villeneuve-d'Ascq	Nord-Ouest	26	0
					<b>Total</b>	<b>33</b>

---

## ANNEXE 4

### Groupements d'intérêt de recherche (GDR), unités d'appui et de la recherche (UAR) et International Research Laboratory (IRL) auxquels la section 34 est rattachée

---

GDR 2195	Langues et Langage à la croisée des Disciplines	LLcD
GDR 2045	Linguistique Informatique, Formelle et de Terrain	LIFT2
GDR 2148	Réseau d'Acquisition des langues secondes	GDR_ReAL2
GDR 2187	Recherches autour des questions d'Éducation	RT CNRS Éducation
IRL 2026	Humanités et sciences sociales - CHICAGO	IRL Chicago
UAR 2004	Maison des Sciences de l'Homme de Bordeaux. Territoires Plurilingues, Sociétés Inclusives, Humanités Évolutives (MSHBx)	MSHBx
UAR 2203	Unité support au Carnot Institut Cognition	Carnot Cognition
UAR 2259	Appui à la Recherche et Diffusion des Savoirs	ARDIS
UAR 2276	POUCHET	POUCHET
UAR 3225	MSH Mondes	MSH Mondes
UAR 3414	MSH et de la Société de Toulouse	MSHS-T
UAR 3598	IR* Huma-Num	IR* Huma-Num
UAR 2503	MSH du Pacifique	MSHP
UAR 3141	Centre Français de Recherche de la Péninsule Arabique	CEFREPA
UAR 3261	MSH Lorraine	MSH Lorraine
UAR 3336	Afrique au Sud du Sahara	
UAR 3337	Amérique Latine	CEMCA
UAR 3565	MSH et de la Société de Poitiers	MSH Poitiers
UAR 3566	MSH et de la Société Sud-Est	MSHS Sud-Est

