

## langues, langage, discours

### *Président*

Christian Hudelot

### *Membres de la section*

Nicholas Asher

Gabriel Bergounioux

Colette Bernard

Claire Beyssade

Isabelle Brill

Eva Buchi

Michel Charolles

Francis Corblin

Bénédicte Guillaume

Sylvie Mellet

Christine Meunier

Lorenza Mondada

Daniele Monseur

Noël Nguyen

Thierry Olive

Martine Vanhove

Sophie Wauquier

Chantal Wionet

La section 34 «Langues, langage, discours» (qui a fait suite à la Section «Sciences du langage» créée au CNRS en 1982) regroupe un champ d'investigations ayant la linguistique comme centre de gravité. Science autonome et diversifiée, la linguistique a pour objectif l'étude scientifique du langage humain à travers l'analyse de la diversité des langues, de leurs propriétés universelles et de leurs propriétés particulières, de la dynamique de leur expansion et de leur diversification diachronique, diatopique et sociale. Cette discipline se situe au carrefour des sciences cognitives et des technologies de l'information puisque le langage et son étude offrent un moyen d'accès à la pensée humaine et aux mécanismes de traitement de l'information permettant d'acquérir, d'utiliser et de transmettre des connaissances.

Dans la mesure où le langage apparaît comme une des capacités spécifiques à l'espèce humaine, mais comme également les langues diffèrent et doivent être apprises dans des groupes sociaux, la linguistique se trouve nécessairement à l'intersection des sciences humaines et sociales, et des sciences de l'esprit. Cette position se manifeste dans le fait que les Unités de recherches qui relèvent de la section 34 tissent des liens non seulement avec d'autres sections de l'INSHS, mais également avec d'autres instituts et principalement l'INS2I et l'INSB. Ainsi, alors que 16 UMR et 1 GDR ont la section 34 comme seule section de rattachement, 5 UMR ont également un lien avec la section 7 et 4 UMR et 1 GDR un lien avec la section 27, tandis qu'une UPR et un GDR associent entre autres les sections 07, 27 et 34. Dans la mesure où la section recrute de jeunes chercheurs dont les projets sont de plus en plus fréquemment à l'interface de la linguistique et du traitement automatique ou des neurosciences intégratives, ce mouvement est appelé à s'amplifier.

On notera qu'au sein même du CNRS les redécoupages successifs n'ont pas encore entraîné un rattachement de toute la recherche centrée sur le langage humain dans la section 34. Sans doute un certain nombre de chercheurs préfèrent-ils encore mener leurs recherches dans le cadre d'équipes dont les centres d'intérêts dominants sont soit des aires linguistiques spécifiques, soit la psychologie, soit l'informatique. Toutefois les évolutions actuelles et la politique volontariste du CNRS en faveur de l'interdisciplinarité incitent les chercheurs à se regrouper en fonction de leur objet d'étude commun et à rejoindre la section 34. Il s'agit néanmoins d'une évolution lente dont le mouvement est loin d'être achevé.

On se doit, pour tenir compte de la place des sciences du langage dans le paysage de la recherche en France, de mentionner une différence notoire qui existe dans la structuration des disciplines entre le CoNRS et la CP-CNU.

Certes, la section 34 du CNRS recoupe principalement la section 07 du CNU Sciences du langage: linguistique et phonétique générales, mais du fait de la structuration de l'enseignement supérieur par grandes aires linguistiques, des linguistes sont également inscrits dans les sections 08 langues et littératures anciennes; 09 langue et littérature françaises; 11 langues et littératures anglaises et anglo-saxonnes; 12 langues et littératures germaniques et scandinaves; 13 langues et littérature slaves; 14 langues et littérature romanes: espagnol, italien, portugais; 15 langues et littératures arabes, chinoises, japonaises, hébraïques; 73 cultures et langues régionales ; etc. Notons que l'on trouve aussi dans la section 27 informatique, des linguistes qui travaillent sur l'interface avec le traitement automatique des langues et dans la section 16 psychologie, des linguistes qui s'intéressent aux questions d'acquisition.

Enfin, il convient de souligner que les recherches françaises dans le domaine des sciences du langage s'appuient sur des collaborations internationales nombreuses et durables, dans le cadre de projets ERC, de projets ANR binationaux, de PICS, etc.

## 1 L'étude du langage humain

La présentation qui suit tente de montrer la dialectique constitutive de la section entre diversité des approches théoriques et méthodologiques et unité de l'objet d'étude: c'est bien le souci constant de comprendre le fonctionnement du langage et des langues qui assure la cohésion de l'ensemble des recherches couvertes par (ou accueillies au sein de) la 34.

### 1.1. Linguistique générale

La linguistique générale a pour but l'étude du langage et des langues dans leur diversité et par delà la diversité des langues. Elle se nourrit d'un aller retour entre le développement de théories et de modèles d'analyse et la mise à jour de données nouvelles dans des langues variées dont l'étude des structures et des propriétés est au cœur de la discipline.

Elle a une interaction forte avec le domaine de la cognition, de la psycholinguistique, du traitement automatique, comme en témoigne le nombre de laboratoires rattachés à la fois aux sections 34 et 27 ou 35; dans ces interactions fructueuses, les linguistes doivent veiller à apporter leur expertise de linguistes.

### Domaines de la linguistique générale

La linguistique générale se subdivise en différents sous-domaines : phonétique, phonologie, morphologie, syntaxe, sémantique, pragmatique, qui s'intéressent aux sons, aux mots, aux phrases ou aux discours.

#### a) Phonétique

La phonétique et la phonologie étudient la forme sonore du langage. Alors que les travaux en phonologie se concentrent davantage sur des aspects systémiques,

la phonétique trouve son champ d'investigation dans l'observation des réalisations physiques et physiologiques des sons. Toutefois, les limites entre les deux disciplines sont de plus en plus poreuses dans la mesure où la phonologie nourrit désormais souvent ses modèles à la lumière de données empiriques et que les analyses phonétiques sont systématiquement mises en perspective avec les structures des systèmes phonologiques et le fonctionnement de la parole dans sa globalité (en lien avec les autres domaines linguistiques comme la morphologie, la syntaxe, le discours ou encore la pragmatique). Ainsi la phonétique, souvent désignée aujourd'hui sous le terme plus large de «sciences de la parole», déploie ses activités dans des champs assez variés nécessitant des collaborations interdisciplinaires (voir section 1.2). Son objectif principal est de comprendre les processus de production et de perception de la parole. Cet objectif est nourri par des recherches concernant des aspects historiques (changement phonétique), cliniques (description des pathologies de la parole), statistiques (exploitation des grands corpus de parole). Ses méthodologies sont le plus souvent expérimentales et les analyses traditionnelles de la parole (acoustiques et articulatoires) sont désormais complétées par les développements récents d'autres disciplines scientifiques (psychologie, neurosciences, traitement automatique).

#### b) Phonologie

Les recherches en phonologie se déploient selon un spectre qui va de l'émergence des structures sonores à partir du traitement du signal –comme en témoigne le courant de la phonologie de laboratoire (Labphon)– jusqu'à des formalismes intégrés à la linguistique cognitive. Entre la connivence avec la phonétique expérimentale et l'ajustement mentaliste, un certain nombre d'écoles, le plus souvent issues du générativisme, témoignent de la vitalité du domaine, que soient interrogés les composants ultimes (phonologie des traits), la constituance prosodique (théorie métrique), les contraintes de combinaison (théorie de l'optimalité) ou l'interaction entre phonologie et morphologie (CVCV). Ces études ont des répercussions importantes pour des secteurs aussi divers que l'acquisition du langage et la linguistique clinique, les travaux sur la graphie ou sur la poésie.

#### c) Morphologie

Les recherches en morphologie sont centrées sur la question de la formation et de la structure des mots, qu'on entende par là les unités constitutives du lexique (morphologie constructionnelle) ou les unités qui constituent les atomes de la syntaxe (morphologie flexionnelle). Elles se sont organisées au cours des dernières années autour de la morphologie lexématique, qui met en doute l'idée selon laquelle les mots sont formés par la simple concaténation d'unités de base appelées morphèmes et soutient au contraire que l'unité de base est le lexème et que les mots sont contruits par application de fonctions à ces lexèmes.

La morphologie constructionnelle est maintenant conçue comme une composante unique, à cause de sa relation

particulière au lexique. Les différents modes de formation (dérivation, composition, conversion) et types de construction de la forme (concaténation, altération radicale, application d'un patron, etc.) sont étudiés en tant que tels au sein de cette composante. En morphologie flexionnelle, la période récente a vu une réanalyse morphologique de nombreux phénomènes qui étaient précédemment conçus comme relevant de la phonologie (allomorphes du radical, classes flexionnelles, syncrétismes, etc.). La question du rôle des paradigmes est redevenue centrale dans les débats.

Deux tendances lourdes traversent l'ensemble du champ : d'une part, l'exploitation systématique de ressources électroniques (corpus, bases de données) et l'approche typologique de la morphologie qui ont radicalement amélioré la finesse et l'empan des descriptions; d'autre part, les travaux à l'interface avec la psycholinguistique, en particulier autour des questions du lexique mental et de l'acquisition des relations morphologiques.

## d) Syntaxe

Les recherches en syntaxe concernent les catégories grammaticales, la structure des phrases et l'ordre des mots. Elles concernent aussi bien la comparaison des langues, l'histoire et l'évolution des langues, que la description des langues en synchronie (en tenant compte de la variation dialectale).

Des recherches nouvelles se déploient sur l'interface entre la syntaxe et la sémantique, la pragmatique ou la prosodie, mais aussi sur la syntaxe des langues rares ou peu connues, et en ce qui concerne les langues européennes, sur la syntaxe des usages oraux ou informels.

Les recherches en syntaxe sont menées dans des cadres théoriques variés : fonctionnel, typologique, grammaires de construction d'inspiration cognitive, grammaire générative dans ses diverses versions, dont les grammaires d'unification (ou grammaires de contraintes).

Ces recherches donnent lieu à la constitution de grammaires de référence et de dictionnaires enrichis d'informations syntaxiques (cf § 4 .1 sur les ressources linguistiques).

## e) Sémantique

Les recherches en sémantique modélisent le sens des mots, des phrases, des textes et des discours. La sémantique lexicale s'intéresse au sens des mots et morphèmes et aux relations de sens entre les mots (synonymie, polysémie, antonymie...). La sémantique compositionnelle s'intéresse à la façon dont on calcule le sens des unités de taille supérieure aux morphèmes ou mots (syntagmes, phrases, textes ou discours). L'idée est que le sens d'un tout est toujours fonction du sens de ses parties et de leurs modes de composition : ainsi par exemple, à partir d'une représentation du sens des mots et de la façon dont ils sont agencés dans une phrase (structure syntaxique), on calcule de façon récursive une représentation du sens de la phrase. On peut distinguer

au moins deux grands courants dans les études en sémantique.

La sémantique d'inspiration cognitive, quant à elle, met l'accent sur la compréhension du langage conçue comme un processus mental. Elle rejette l'idée d'une autonomie du langage par rapport à la cognition, postule l'existence d'invariants non linguistiques et met en avant les notions d'élaboration mentale et de représentation mentale des significations.

Pour représenter la signification des mots et des textes, il est fréquemment fait usage de réseaux sémantiques inspirés de la théorie des graphes en mathématiques, et de « frames ». Ces dernières années, de nombreux travaux ont été menés, à la fois par des linguistes, des philosophes et des psychologues sur la cognition située: les linguistes cognitivistes affirment en effet que le langage est toujours à la fois « incarné » et « situé » dans un environnement spécifique.

La sémantique vériconditionnelle assimile le sens d'une phrase à ses conditions de vérité. Dans les dernières décennies, les travaux menés sur la sémantique des temps verbaux, des indéfinis, des anaphores et des présuppositions ont suscité l'émergence de théories sémantiques dites dynamiques qui visent à rendre compte du flux informationnel lié à la dynamique du discours : la signification linguistique est conçue comme une fonction qui modifie le contexte discursif, et ne s'analyse pas uniquement en termes d'impact sur les conditions de vérité de l'énoncé. Dans la période récente, les recherches ont porté à la fois sur des questions explorées depuis longtemps comme l'étude des expressions quantifiées et de choix libre (tel *n'importe quel*) ou l'analyse des modalités et sur des champs nouvellement ouverts comme l'analyse du sens des énoncés non assertifs (questions, ordres...) ou la question de la répartition du travail entre sémantique et pragmatique (dans le traitement des implicatures et des présuppositions au travers de la question des différents niveaux de sens ou dans l'analyse des relations rhétoriques).

La sémantique dynamique a étendu les méthodes de la sémantique compositionnelle au-delà de la phrase jusqu'au traitement des discours et même des dialogues. Il y a des interactions importantes entre la sémantique dynamique et les développements en logique et en informatique théorique: des chercheurs français ont montré par exemple comment enchâsser la sémantique dynamique dans le cadre de la logique du deuxième ordre de Montague en utilisant la méthode des continuations de l'informatique théorique.

## f) Pragmatique

Les travaux en pragmatique s'occupent des enrichissements de la sémantique qui sont dus à la prise en considération du contexte d'énonciation. La frontière entre la sémantique et la pragmatique est un sujet de débat dans la philosophie contemporaine du langage et beaucoup de linguistes considèrent que la frontière entre les deux domaines est flexible. La période récente a continué à interroger ces frontières en s'orientant

vers une pragmatique de type formel et en discutant la manière dont la valeur pragmatique des items particuliers se combine dans l'interprétation des phrases complexes. Parallèlement, des travaux récents à l'interface de la linguistique et de la psychologie expérimentale ont entrepris de mettre à l'épreuve un certain nombre de propositions théoriques concurrentes. Ces travaux portent en particulier sur les présuppositions, les implicatures scalaires et la contribution au sens des relations discursives et de la structure discursive d'un texte ou d'un dialogue.

La pragmatique s'est développée en interrogeant de manière féconde ses interfaces avec la syntaxe, la prosodie et, plus récemment, avec la gestualité et la multimodalité, montrant l'importance d'approches multi-dimensionnelles et multi-fonctionnelles, et d'une conception du langage qui intègre pleinement ses dimensions situées et contextuelles. De manière plus générale, la pragmatique s'est élargie aux analyses du discours et à la linguistique de l'interaction, qui ont permis une étude des structures de la langue telles qu'elles émergent et se développent dans ses usages situés dans le dialogue, la conversation et des situations de communication spécifiques (cf § 2.5).

La pragmatique, tant dans sa dimension discursive que dans sa dimension formelle est bien représentée en France.

## g) Texte & Discours

Les linguistes travaillant sur les textes et les discours se doivent de rassembler, structurer et décrire leurs objets empiriques pour les constituer en observables, c'est-à-dire en objets scientifiques construits sur la base desquelles ils peuvent élaborer des hypothèses explicatives et des modélisations susceptibles de représenter leurs fonctionnements et de fournir par là même les bases d'un dialogue fructueux notamment avec les sciences cognitives. Une telle construction oblige à définir les unités textuelles et discursives élémentaires, à penser leur articulation et à appréhender leur combinatoire (nécessairement complexe puisque ces unités sont de différents niveaux : lexical, syntaxique, sémantique, discursif, etc. et que ces niveaux se superposent, s'imbriquent), à enrichir les corpus et à se doter des outils logiciels et des méthodologies nécessaires à leur traitement (on rejoint ici certains aspects du TAL). Un tel outillage est aujourd'hui indispensable pour appréhender les structures informationnelles et rhétoriques des textes et des discours.

Les modèles spécifiquement dédiés au discours, comme ultime palier d'organisation du langage, ont une couverture plus ou moins large. Les modèles les plus formels (comme la 'Rhetorical Structure Theory') reposent sur un répertoire de relations de discours dont l'affectation s'effectue au fur et à mesure de l'avancée du texte. Ils peuvent également (comme la 'Segmented Discourse Representation Theory') prendre en compte les relations entre les référents (entités, propositions et situations) et inclure des procédures de mise à jour. Parallèlement à ces modèles, un grand nombre d'analyses linguistiques locales (fondées sur des corpus de langue écrite ou parlée) qui ne couvrent pas l'ensemble des niveaux de structuration des discours, portent sur des items comme

les anaphores et les connecteurs, dont l'interprétation ne peut être établie indépendamment de ce qui a été dit ou écrit précédemment, ou sur des expressions fixant des repères pour l'interprétation de ce qui va suivre.

L'analyse du discours a été en outre fortement renouvelée par l'analyse de la langue orale en interaction, proposée notamment par la linguistique interactionnelle inspirée de l'analyse conversationnelle, la linguistique fonctionnelle et la grammaire de construction. L'enjeu est de montrer le lien entre l'organisation de la grammaire et des ressources linguistiques, l'organisation des environnements séquentiels dans lesquels elles sont utilisées en interaction et l'organisation de l'action pratique qu'elles accomplissent. Cela donne lieu à des analyses de la syntaxe, de la prosodie, du lexique, des particules discursives telles qu'elles se déploient dans des contextes d'usage précis. Cela permet aussi une meilleure connaissance de l'organisation interactionnelle, domaine d'émergence du changement linguistique.

## Modélisation

Les théories linguistiques, comme celles issues des grammaires génératives ou des grammaires catégorielles, peuvent être formalisées. Les recherches en linguistique formelle se caractérisent par une grande variété de modèles, dont les propriétés formelles peuvent être étudiées en tant que telles. La sémantique formelle par exemple utilise comme langage de représentation le langage de la logique, du premier ou du deuxième ordre. En France, des chercheurs de renommée internationale travaillent dans plusieurs laboratoires du CNRS sur la question du pouvoir expressif des langages formels utilisés pour modéliser différentes propriétés des langues naturelles.

Cette activité est importante de plusieurs points de vue. Les fondements formels d'un modèle linguistique peuvent être évalués en termes de complexité ou de vraisemblance cognitive, et comparés à d'autres modèles linguistiques, et plus généralement à d'autres systèmes cognitifs. Les langages de formalisation peuvent permettre la réalisation de programmes informatiques qui implémentent le modèle, et sont utiles en traitement automatique des langues (voir partie 4).

## Histoire des théories linguistiques

La connaissance des travaux antérieurs est une condition des avancées de la discipline dans une perspective cumulative. Comme l'a dit Bachelard, une science est condamnée à répéter les erreurs dont elle n'a pas gardé la mémoire. L'histoire de la linguistique est sollicitée dans trois dimensions. La première, descriptive, entend corriger, éditer et présenter les travaux passés dans une démarche documentaire qui s'apparente à celle du philologue: les « humanités digitales », en cours de développement, en sont une illustration. La seconde, critique, propose une évaluation des analyses produites à l'aune de celles qui ont déjà été faites, que cette antériorité remonte à des siècles ou à quelques années, comme il en va des discussions continues, depuis Aristote jusqu'aux

philosophies du langage ordinaire, concernant le type de relation à établir entre logique et langage. La troisième, heuristique, entre dans le panorama plus général de l'épistémologie en décrivant dans quelles conditions intervient et se sanctionne une découverte dans les sciences du langage.

## 1.2. Les sciences de la parole

La parole est un sujet d'étude fondamentalement pluridisciplinaire dans la mesure où il concerne à la fois des aspects linguistiques (spécificités des systèmes linguistiques), des aspects fonctionnels du corps humain (implication des organes de la parole, mais aussi des yeux, des mains), des aspects cognitifs (programmation, perception), et des aspects physiques (signal de parole). Cette pluridisciplinarité est depuis longtemps structurante pour les recherches sur la parole et elle continuera à l'être dans les années à venir.

Un axe de recherche très actuel vise à établir des liens entre génétique, évolution des systèmes phonologiques et ontogenèse de la parole. On cherche par exemple à caractériser avec précision les contraintes articulatoires, acoustiques, perceptives et cognitives susceptibles de s'exercer sur l'architecture des systèmes phonologiques, grâce à des études expérimentales à cheval entre la phonologie, la phonétique et la psychologie cognitive. Ces recherches ont parfois pour objectif de reproduire en laboratoire les phénomènes dont on suppose la mise en jeu dans l'évolution des systèmes phonologiques. Il est par ailleurs largement fait recours aux modèles computationnels (modèles connexionnistes, systèmes artificiels multi-agents, etc.) dans le but de simuler la manière dont les unités phonologiques peuvent émerger dans la mise en relation entre son et sens.

Depuis quelques années, les collaborations entre phonéticiens et médecins cliniciens se sont renforcées dans le but d'aboutir à une description plus objective et à grande échelle de la parole pathologique. Ces travaux visent aussi à mieux cerner les mécanismes cérébraux en jeu dans la production et le traitement de la parole chez le sujet normal à la lumière des études réalisées sur des sujets cérébro-lésés. Ces collaborations, fructueuses, sont appelées à s'étendre davantage encore.

La prosodie a été longtemps négligée ou envisagée comme un composant marginal du champ de la linguistique, voire exclue de ce dernier. Le développement considérable des recherches sur la prosodie au cours des dernières décennies est ainsi à la fois le reflet et l'un des moteurs de l'évolution des paradigmes de la linguistique contemporaine. L'avènement de la phonologie plurilinéaire, l'extension du champ d'investigation de la linguistique aux usages langagiers dans le dialogue et la conversation quotidienne, l'intérêt porté au rôle de la prosodie dans l'acquisition du langage par l'enfant, forment quelques-uns des éléments dans cette évolution qui se poursuit aujourd'hui au travers de multiples travaux.

La parole est une activité qui implique le corps humain au travers des organes de la parole (langue, lèvres, larynx, etc.) mais aussi au travers de l'ensemble des gestes

qui accompagnent et structurent la parole comme les mouvements des sourcils, ceux des mains, la posture, etc. De nombreuses recherches se développent à la fois sur l'extraction du sens et la codification des gestes en lien avec la parole, mais aussi sur la structuration des discours et des interactions dans des situations de parole naturelle.

Ces recherches vont de pair avec un courant bien engagé en phonétique depuis une décennie et qui consiste à observer les caractéristiques de la parole sur de très larges corpus de parole naturelle (non contrainte). Les outils automatiques de traitement de la parole permettent désormais d'envisager l'exploitation de très gros corpus pour élargir nos connaissances sur les contraintes de production et de perception de la parole en condition naturelle, c'est-à-dire avec l'ensemble des facteurs liés à la communication (sources de variations de la parole).

Il existe entre les recherches sur la forme sonore du langage et le traitement automatique des langues (TAL) des connexions bien établies depuis longtemps. La construction des systèmes de synthèse et de reconnaissance automatiques de la parole en a été l'un des moteurs. Au cours de ces dernières années, ces relations se sont renforcées sous l'influence de différents facteurs. Le premier point de convergence tient aux techniques de recueil et d'exploitation partagée des ressources considérables dont nous disposons à présent dans le domaine de l'oral. Le second point de convergence est lié à l'intérêt croissant des spécialistes du TAL pour les langues parlées. Dans les recherches sur l'oral, on voit ainsi se généraliser l'utilisation d'un large ensemble d'instruments et de méthodes permettant de procéder de manière automatique au traitement et à l'analyse des données recueillies. Les études sur corpus de grande ou très grande taille menées en phonologie et en phonétique, font ainsi de plus en plus souvent appel à des outils initialement développés en TAL. Les spécialistes du TAL et les phonologues/phonéticiens ont également contribué de manière conjointe au développement des systèmes d'annotation multi-niveaux.

## 2. Etudes des langues et des discours

### 2.1 Etude des langues dans leur diversité

L'étude et la description des langues dans leur diversité et sous tous leurs aspects (phonique, morpho-syntaxique, lexical, sémantique) constitue un des pôles essentiels de la discipline et a des implications pour la typologie des langues et les sciences cognitives.

Outre les connaissances nouvelles qu'elles apportent sur les familles de langues, leur histoire et leur comparaison, les données collectées permettent l'émergence de problématiques nouvelles en mettant les théories à l'épreuve des faits: il n'est pas rare que l'on découvre des données qui falsifient les prédictions des théories et qui obligent à les réviser.

La constitution de grammaires et de dictionnaires est d'une utilité fondamentale pour les chercheurs comme pour les locuteurs des langues. Cette documentation est la base indispensable des recherches typologiques et comparatives. Elle permet d'enregistrer des états de langue, clé essentielle pour leur comparaison et pour la compréhension de leur évolution. Les dictionnaires mettent à la disposition des chercheurs des bases lexicales indispensables à toute entreprise de comparatisme historique et de reconstitution du vocabulaire de base commun à une famille donnée.

Enfin, la mise en œuvre de systèmes d'écriture dans des langues à tradition orale, les dictionnaires ou atlas encyclopédiques permettent aussi de sauvegarder la diversité du patrimoine linguistique et culturel et se font en étroite relation avec le domaine très dynamique de l'anthropologie linguistique (Hymes, Duranti) et de l'anthropologie-ethnologie en général.

La France a une longue tradition de recherches qui concerne aussi bien les langues anciennes, le français dans sa diversité, et les langues du monde. Deux fédérations couvrent ces différents domaines, l'ILF (Institut de Linguistique Française), et la Fédération T.U.L (Typologie et Universaux du Langage) et plusieurs laboratoires du CNRS se consacrent à cette tâche: citons à titre d'exemple les langues du Caucase, tibéto-birmanes, africaines, amérindiennes ou de la région Asie-Pacifique.

Le CNRS a une mission plus particulière concernant les langues autochtones d'Outremer, faisant partie des « Langues de France » et pour lesquelles les travaux des linguistes français ont des applications immédiates dans l'enseignement primaire, secondaire et universitaire.

Le départ à la retraite dans les cinq prochaines années d'un grand nombre de spécialistes de renom risque, si une politique de recrutement de jeunes chercheurs n'est pas mise en place, de mettre à mal des domaines dans lesquels la recherche française s'est distinguée et est reconnue au plan international.

## 2.2 Typologie linguistique

En dépit de la grande diversité des langues du monde, des traits et des phénomènes structurels et fonctionnels communs, parfois appelés universaux (mais ce sont surtout des universaux statistiques, les universaux absolus étant rares), ont été recensés.

Des traits et phénomènes relevant de divers sous-domaines (phonologie, morpho-syntaxe et sémantique) s'organisent en types dont on analyse la distribution et la co-occurrence avec d'autres phénomènes pour mettre en évidence des corrélats et dégager des implications ou des universaux implicationnels : par exemple, si une langue a des phrases à verbe final elle a normalement des postpositions.

Parmi les domaines d'investigation des typologues, citons notamment la typologie tonale, l'ordre des mots, les structures et marques d'actance et types d'alignement morpho-syntaxique, le type de dépendance avec marquage sur la tête ou sur le terme dépendant, la constitution de

cartes ou arbres sémantiques de traits grammaticaux ou lexicaux, etc.

Les types dégagés constituent des critères de classification des langues, qui ne sont dus ni à des liens génétiques, ni à des faits historiques, ni à des faits aréaux (de contact), mais qui relèvent davantage de phénomènes cognitifs généraux exprimés à travers la diversité des langues, et de la manière dont le cerveau humain traite et organise l'information. Des travaux récents partent du constat que les universaux du langage sont peu nombreux, et voient au contraire la diversité linguistique comme le point crucial pour les sciences cognitives. Un dialogue fructueux peut alors s'engager avec certains champs de recherche en biologie, centrés sur la problématique de la diversité et de la variation.

Pour avoir une certaine validité, ces recherches sont menées extensivement sur un grand nombre de langues typologiquement différentes ; elles s'articulent à de grandes bases de données dont le traitement relève de la typologie quantitative.

Des recherches actuelles tentent également de dégager des universaux sémantiques statistiques et un cœur de notions communes à toutes les langues, concernant par exemple la terminologie des parties du corps, des émotions, etc.

Depuis une vingtaine d'années, des chercheurs français se sont engagés dans la typologie et la quête d'invariants du langage. Ces recherches ont pris un essor important ces dix dernières années grâce à la création de la fédération T.U.L, qui a été le catalyseur de très nombreuses recherches conduites dans des approches théoriques différentes. Les recherches menées au sein de la Fédération T.U.L ont un triple but :

(i) la définition des propriétés communes aux langues (universaux) et des contraintes générales auxquelles elles sont soumises à tous les niveaux structurels, que ces contraintes résultent de leur architecture interne ou de contraintes cognitives générales ;

(ii) articuler ces recherches à l'analyse de la variation inter-langues et de la diversité linguistique, afin de dégager les mécanismes internes et externes d'évolution des systèmes, et leur modélisation ;

(iii) élaborer des bases de données linguistiques et typologiques, contribuer à l'archivage de textes de langues en danger ou peu documentées.

La fédération T.U.L a également favorisé de nombreux liens institutionnels internationaux, comme en témoignent l'organisation, pour la première fois en France, de la 7ème édition de la conférence internationale de l'ALT (Association of Linguistic Typology) et celle d'une Ecole d'été franco-allemande.

## 2.3 Linguistique historique, classification et évolution des langues

L'étude des langues dans leur dimension historique, évolutive et leur classification en familles est un domaine fondamental de la linguistique.

Au cours des dernières années les linguistes ont continué à avancer dans le déchiffrement et l'interprétation des langues littéraires éteintes (comme le méroïtique), la reconstruction des proto-langues, la compréhension de l'histoire des familles de langues, de leur structure interne comme de leurs parentés lointaines, et des contacts qui les ont influencées. L'évaluation de ces propositions se poursuit. Dans de nombreux domaines un certain scepticisme face aux propositions de parentés lointaines faites dans la période précédente semble prévaloir. Cependant la reconnaissance toute récente par de nombreux linguistes d'une parenté génétique entre une langue du vieux monde, le ket, parlé dans la vallée de l'enissei, avec les langues athabasques du continent américain, est un événement particulièrement saillant.

Un renouveau méthodologique est en cours. L'utilisation des arguments typologiques dans la détermination des parentés linguistiques se répand. L'application des statistiques bayésiennes à la distribution des groupes de cognats dans les langues d'une famille ouvre des perspectives intéressantes pour l'élucidation de sa phylogénie, comme d'ailleurs d'autres méthodes importées des sciences de l'évolution telles que les méthodes de compatibilité appliquées au vocabulaire de base. D'autres travaux suggèrent que la longévité des mots est corrélée à leur fréquence d'emploi. L'étude systématique des changements sémantiques attestés aide à mieux saisir leur directionnalité, tout comme l'expérience accumulée par les études sur la grammaticalisation. La susceptibilité des mots à l'emprunt a fait l'objet d'études quantifiées dont les résultats pourraient conduire à la définition d'une liste de vocabulaire de base établie sur des bases explicites, à la différence des listes de type Swadesh.

Alors que l'idée d'un peuplement du monde par notre espèce au cours des derniers 150 000 ans semble maintenant largement acceptée, la capacité de linguistes à proposer des modèles de peuplement leur permet de participer aux discussions avec les généticiens des populations, archéologues, paléo-climatologues, etc. portant sur la période de formation des grandes familles de langues modernes au cours des 10.000 dernières années.

Enfin, les développements de nouvelles théories sur les contacts de langues, la constitution et l'évolution des pidgins et créoles, le changement linguistique, la grammaticalisation, les alternances codiques ont aussi contribué à une profonde évolution de la linguistique historique et comparative. Cette évolution se poursuit.

## 2.4 Sociolinguistique, diversité des pratiques langagières, variation

La diversité des langues et de leurs usages est au cœur

de la discipline et demande une prise en compte, à côté de la diversité des systèmes linguistiques, de la variation des formes et des pratiques en contextes. C'est ainsi que les recherches relevant de la dialectologie, de la sociolinguistique, de l'anthropologie linguistique, de la linguistique interactionnelle contribuent à la documentation, description et modélisation de la diversité et de la variation au sein d'aires linguistiques, de groupes sociaux, de contextes d'activité. Les enjeux se situent aussi bien sur le terrain, en concernant la multiplicité des techniques d'enquête et de documentation – des questionnaires et des cartes aux recueils d'enregistrements en situation non contrainte par le chercheur – que du côté de la conceptualisation des phénomènes de variation, de changement, d'émergence, d'ajustement contextuel. Les corpus de données documentant les langues, leurs variétés et leurs usages en situations diversifiées doivent continuer à se développer afin de mettre à disposition des données 'écologiques' en mesure de rendre compte des dynamiques des non-standards, des alternances de styles, des accommodations linguistiques – que ce soit dans des domaines aussi divers que, par exemple, les pratiques linguistiques dans des domaines d'expertise professionnelle, dans des groupes de jeunes, dans des communautés migrantes, dans les interactions médiées par les nouvelles technologies. Un domaine particulièrement important est représenté par les situations de contact, caractérisées par des alternances codiques mais aussi des *lingua francae* hétérogènes et des mélanges de langue, produisant des variétés hybrides et émergentes. Ces hybrides linguistiques mettent à l'épreuve les modèles basés sur une conception homogène de la langue et invitent à de nouvelles conceptualisations. Cela met au cœur de l'étude de la diversité les processus de catégorisation à la fois des langues, des locuteurs, des contextes de parole et d'interaction, des activités et des relations mutuellement configurantes entre ces catégorisations et les conduites effectives des locuteurs.

## 2.5 Textes, discours et interaction

L'essor de la linguistique de corpus durant la période récente s'est traduit par la mise à disposition de vastes ressources et de systèmes d'annotation de plus en plus raffinés. Les données désormais accessibles ouvrent des perspectives nouvelles pour le développement des études sur les textes et les discours.

L'analyse de la textualité à proprement parler ouvre un champ original et en plein essor. Les modèles sont en partie contraints, et par la langue, et par des patrons de production, dont les genres sont un exemple largement reconnu mais encore mal exploré. Comprendre la mise en œuvre de cette complexité (tant au niveau de l'encodage que du décodage) et tenter de la modéliser conduit, là encore, à des collaborations étroites avec l'informatique, les mathématiques (statistiques et topologie) et la psychologie expérimentale. A ces travaux, viennent s'ajouter des études explorant les conditions historiques d'emploi de la langue, dans leur dimension à la fois textuelle et sociale. Ces études mettent l'accent soit sur l'interdiscours où s'entrecroisent des séries d'énoncés dont on peut étudier les récurrences et les déplacements, soit sur des pratiques discursives dont les fonctionnements ne se ramènent

pas seulement à des contraintes pragmatiques ou à des mécanismes rhétoriques universels, et dont il faut restituer les conditions historiques.

Le développement des recherches sur le texte et le discours doit être favorisé dans la mesure où les résultats acquis et attendus ont une incidence directe sur les travaux

- en linguistique générale et comparative, pour l'étude des phénomènes de grammaticalisation (les régularités pragmatiques que l'on observe en discours tendant, avec le temps, à devenir des contraintes syntaxiques) ;
- en TAL, pour la hiérarchisation des informations et donc le filtrage, ou pour la segmentation thématique ;
- en psycholinguistique, pour le développement et l'évaluation d'hypothèses sur les processus de production et de compréhension afin de mieux comprendre l'acquisition des habiletés et compétences textuelles mais aussi les diverses pratiques de production.

Ce développement devrait être accompagné par la mise à disposition des unités de moyens en personnels et en équipements leur permettant de mettre en œuvre des études sur de très larges corpus (gros travail d'annotation et de traitement des données) et d'étoffer leurs forces en TAL et en psycholinguistique.

### 3. Langage et Cognition

#### 3.1 philosophie du langage

Les recherches en philosophie du langage ont des liens importants avec la sémantique, la pragmatique et les fondements de la linguistique théorique. Longtemps pratiquée presque exclusivement dans le monde anglo-saxon, la philosophie du langage est un domaine maintenant bien représenté et bien vivant en France. Il y a des collaborations régulières entre les chercheurs de haut niveau américains et britanniques (venant notamment du MIT, de Yale, Harvard, Rutgers, UCLA, Oxford ou Cambridge) et plusieurs laboratoires du CNRS, grandes écoles et instituts de recherche français (INRIA, EHESS, ENS). La communauté des chercheurs travaillant sur le sens et l'interprétation tant du point de vue linguistique que logique ou philosophique s'est structurée ces dernières années autour du GDR de sémantique et a acquis une meilleure visibilité à la fois à l'échelle nationale et internationale, comme en témoigne l'organisation de conférences annuelles en France (comme SPE ou JSM). Les recherches en France sur les processus mentaux qui sous-tendent le traitement du langage (l'acquisition, la production, la perception et la compréhension) et les liens qui s'établissent dans ces processus entre langage et pensée en ont grandement profité.

#### 3.2 psycholinguistique

La psycholinguistique vise à comprendre les différents facteurs psychologiques et neurophysiologiques qui rendent l'être humain capable d'acquérir et d'utiliser le langage. Plus précisément, cette discipline cherche à décrire les processus mentaux qui sous-tendent la production et la compréhension du langage, du point de

vue de l'acquisition et de l'utilisation experte. Ainsi, de nombreux psycholinguistes cherchent à définir la part de l'universel par rapport aux spécificités des langues, analysent les interactions entre le développement cognitif et le développement langagier, ou bien les interactions entre la conceptualisation d'un message à produire, et la formulation de ce message dans une langue donnée. On cherche aussi à mieux comprendre les origines psychologiques ou neuropsychologiques de différents dysfonctionnements (aphasies, dysphasies, etc.) ou retards d'acquisition (dyslexie, dysorthographe, illettrisme, etc.).

Comprendre comment l'être humain utilise le langage implique d'avoir recours à une diversité d'approches et de disciplines : de la linguistique formelle à la linguistique de corpus, aux paradigmes de la psychologie expérimentale ainsi qu'aux méthodes d'imagerie utilisées dans les neurosciences cognitives. On voit bien ici la nécessaire interdisciplinarité des études sur le langage. Une interaction constante entre sciences du langage et psycholinguistique doit être maintenue. L'observation des productions linguistiques aussi « complètes » et « naturelles » soient-elles ne fournit, en effet, aucune indication sur les processus sous-jacents à la production et à la compréhension. En revanche, les expérimentations psycholinguistiques sont spécifiquement conçues pour pister ces processus, en particulier des indicateurs sur leur déroulement temporel et les structures cérébrales impliquées par ces traitements. De même, la psycholinguistique doit intégrer les apports de la linguistique pour comprendre comment les contraintes liées aux règles et aux structures du langage affectent les traitements de production et de compréhension.

Un exemple qui illustre cette interdisciplinarité est l'étude des langues des signes (LS). Les conditions d'accès des sourds à leurs langues offrent à la psycholinguistique et à la linguistique un ensemble de questionnements féconds. L'acquisition des LS fournit en effet, en soi, un contrepoint important parce qu'il s'agit de langues visuo-gestuelles, présentant de fortes particularités structurelles, voire typologiques (iconicité, multilinéarité, utilisation pertinente de l'espace). En termes développementaux, elles ré-interrogent notamment les liens entre cognition, imagerie mentale et conceptualisation et éclairent, au-delà des seules LS, la question des relations ontogénétiques entre langue et gestualité. Dans la mesure où ces langues sont exceptionnellement langues maternelles (95% des enfants sourds naissent en milieu entendant), elles conduisent aussi à questionner la notion d'âge critique. Il en va de même de la problématique des modalités de l'accès à l'écrit de la langue vocale de personnes qui n'entendent pas, à même d'apporter un éclairage nouveau sur les processus cognitifs en jeu dans l'entrée dans l'écrit comme sur les processus d'acquisition d'une L2.

Constatant cela, il faut prendre acte du fait que, pour l'étude de certains phénomènes linguistiques, un nombre croissant d'unités de la section 34 vont demander à disposer de personnels et d'équipements permettant de conduire des investigations alliant des études sur corpus, des modélisations informatiques, et des études expérimentales.

## 4. Traitement automatique des langues, outils et ressources

Les questions théoriques que se posent les linguistes se nourrissent de la confrontation avec les données authentiques, ce qui incite à la production de ressources linguistiques (bases de données, corpus, dictionnaires). Les sciences et technologies de l'information constituent, avec la cognition, l'autre grand domaine d'interface de la linguistique, sous le nom de traitement automatique des langues (TAL). De nombreux laboratoires d'informatique ont un rattachement secondaire à la section 34, et un certain nombre de chercheurs changent avec profit de section dans un sens ou dans l'autre. Le rôle du CNRS est aussi de recruter dans ce domaine les ingénieurs indispensables à ce type de recherches.

### 4.1 Constitution de ressources linguistiques

Les ressources langagières « brutes » (corpus de journaux ou de radio disponibles sur internet par exemple) sont des matériaux importants pour les linguistes. Mais elles sont insuffisantes quand il s'agit de mener des recherches un tant soit peu précises (Dans quels contextes trouvez-vous un sujet inversé ? Quels adjectifs épithètes précèdent le plus souvent le nom ?). C'est pourquoi une part importante du travail des linguistes est constituée par la construction de véritables ressources linguistiques, à grande échelle et librement disponibles, aussi bien pour les chercheurs que pour le grand public. On peut citer ainsi, à titre d'exemple, la base textuelle Frantext ([www.atilf.fr/frantext](http://www.atilf.fr/frantext)), avec plus de 220 millions d'occurrences de mots, le TLFi ([www.tlfi.fr](http://www.tlfi.fr)) ainsi que le corpus arboré pour le français (*French treebank*), qui ont profondément modifié le travail sur la langue française.

La synthèse des connaissances grammaticales sur des langues variées, de façon systématique et indépendamment des cadres théoriques, est également un objectif à encourager. On a vu paraître ces dernières années d'excellentes grammaires de référence de plusieurs langues européennes comme l'anglais, le catalan, l'espagnol ou encore l'italien, mais pas pour le français. La réalisation en France, dans des laboratoires du CNRS, de grammaires de ce type, l'une pour le français contemporain, l'autre pour l'histoire du français, sont des projets de longue haleine, réunissant plusieurs dizaines de chercheurs français et étrangers, et à encourager. Une grammaire de référence est un ouvrage « visible » et dont la vie utile s'étend sur un nombre appréciable d'années. Elle engage donc aussi l'utilité sociale de la linguistique.

### 4.2 Lexicographie

La lexicographie est le mode d'exposition privilégié de la lexicologie : il s'agit essentiellement d'un langage. En tant que tel, elle vise à modéliser le lexique, la composante la moins aisément saisissable, d'un système linguistique, mettant à nu ses structures sémantiques, morphologiques, syntaxiques ou encore étymologiques.

Contrairement à la lexicographie commerciale, la lexicographie scientifique a appris à renoncer à la course à la nomenclature la plus ample au profit de la qualité de la description. Plutôt que par tranches alphabétiques successives, la sélection de la nomenclature est établie sur une base variationnelle (états de langue du passé, vocabulaire marqué diatopiquement), étymologique (lexique héréditaire, emprunts à une langue donnée) ou thématique (vocabulaire d'un domaine notionnel). La description est réalisée à partir de données textuelles (bases de données, éditions de textes) ou orales (enquêtes de terrain), les exemples construits étant en principe proscrits. Les microstructures adoptées peuvent facilement dépasser la dizaine de niveaux d'analyse et la centaine de marqueurs descriptifs. De façon caractéristique, les produits lexicographiques de pointe peuvent aller jusqu'à se présenter comme des recueils d'articles de revue scientifique classés par ordre alphabétique.

Si la lexicographie constitue un secteur en plein essor, le défi de la recherche française dans ce domaine consiste en une homogénéisation des comportements de recherche et des pratiques d'exposition : l'excellence reste pour l'essentiel confinée aux grandes langues véhiculaires, aux variétés hautes, au lexique non construit, tandis que les langues vernaculaires rares, les variétés diasystématiques peu valorisées ou plus difficiles à appréhender, le lexique morphologiquement construit et les unités polylexicales représentent un terrain encore largement à conquérir. En outre, les principes microstructurels ambitieux tels que l'exigence d'une contextualisation ou celle de l'authenticité des exemples n'ont percé que récemment – et de façon très ponctuelle – dans le domaine de la lexicographie bi et multilingue, qui mérite d'être développée dans le contexte actuel d'une mondialisation croissante.

L'informatisation des productions lexicographiques est à l'heure actuelle presque généralisée, ce qui permet de nouvelles formes de production, mais aussi d'exploitation : le dictionnaire devient un corpus, une ressource exploitable par des sous-disciplines linguistiques comme la morphologie constructionnelle, la microsyntaxe ou la sémantique lexicale, tout en ouvrant des perspectives particulièrement fécondes à la recherche en traitement automatique des langues. En 2005, le CNRS s'est ainsi doté du CNRTL (Centre National de Ressources Textuelles et Lexicales, (<http://www.cnrtl.fr>), dont le Portail lexical permet la consultation croisée et l'exportation dans un format normalisé d'une multitude de lexiques et de dictionnaires.

### 4.3 Linguistique de corpus

La disponibilité de ressources textuelles électroniques de grande taille (corpus, bases de données textuelles, dictionnaires et lexiques) et les progrès de l'informatique ont permis au linguiste d'aller au-delà de l'accumulation de faits de langue et de confronter ses théories à l'usage effectif de la langue. C'est ce qu'on appelle la linguistique « de corpus », qui se développe depuis une quinzaine d'années en France et dans le monde.

La linguistique de corpus est confrontée à de nouveaux

défis et de nouvelles complexités lors du traitement de corpus de langue orale et de langue en interaction. Ces défis sont multiples et traversent toutes les étapes de constitution et d'analyse des corpus oraux : arpentage du terrain en vue de recueillir des données représentatives et de qualité ; enregistrements audio et de plus en plus vidéo en vue de garantir une observabilité maximale des détails non seulement linguistiques mais aussi visuels et multimodaux de l'interaction ; anonymisation et protection de la vie privée des informateurs ; transcription et annotation et leurs standardisations ; mise au point de moteurs de recherche et d'étiquetages automatiques en mesure de traiter le caractère non standard de l'oral et de prendre en compte les dynamiques interactionnelles. Les avancées dans ce domaine permettent aujourd'hui d'envisager des banques de données de corpus qui documentent la langue parlée en milieu naturel, dans des interactions sociales situées dans une diversité de contextes, en tenant compte non seulement de la parole ainsi produite, mais aussi des conduites multimodales (regards, gestes, postures du corps, mouvements etc.). On citera par exemple le projet PFC (Phonologie du Français Contemporain), la plateforme CLAPI (Corpus de Langue Parlée en Interaction), spécialisée en corpus authentiques d'interactions, ou encore le Centre de Ressources pour la Description de l'Oral (<http://www.crdo.fr>), qui héberge de nombreuses ressources provenant de différents laboratoires. Ces données sont encore rares, et les corpus continuent à être de taille inférieure à ceux dont on dispose pour l'écrit – du fait de la difficulté de recueillir de la parole vive et du fait des coûts très élevés de transcription. Une priorité est donc de multiplier les corpus accessibles dans des formats standardisés et traitables par des outils informatiques. L'accès à de grandes masses de données orales et interactionnelles permet des descriptions et modélisations inédites de l'oral, ainsi que des dynamiques d'émergence et de changement de langue et de leurs variations dans des contextes d'usages spécifiques.

## 4.4 Archivage des langues

La constitution d'archives de corpus et de ressources linguistiques de première main, collectées par des spécialistes contribuent à deux grands domaines : d'une part, lorsqu'elles portent sur des langues minoritaires souvent peu décrites, voire inconnues, elles sont l'un des moyens essentiels permettant de documenter et de préserver des langues en danger et de préserver le patrimoine culturel qu'elles véhiculent. D'autre part, lorsqu'elles concernent le français dans ses variations, elles permettent d'alimenter le projet d'un corpus de référence nationale. Les deux Fédérations de linguistique se sont fortement engagées sur ces terrains, TUL pour les langues peu décrites, l'ILF pour les corpus de français.

Le développement d'archives linguistiques étiquetées ou annotées de corpus oraux et écrits est un secteur dont l'essor va croissant au niveau international. Ils constituent des bases de données qui enregistrent un état de langue donné et qui sont essentielles pour les recherches actuelles et futures.

Des avancées notables, avec la mise en ligne de nombreuses données de langues diverses, ont déjà eu

lieu dans ce domaine avec la création des archives du LACITO (<http://lacito.vjf.cnrs.fr/archivage/presentation.htm>), maintenant également hébergées par le CRDO (<http://www.crdo.fr>). D'autres corpus sont en cours de constitution grâce notamment au soutien de projets ANR ou européens. La constitution d'archives et de corpus numériques doit être soutenue.

## 4.5 Traitement automatique des langues

Si la linguistique informatique a connu un grand essor dans les années 1950, ce domaine bénéficie aujourd'hui des apports du traitement de la parole, de la recherche d'information, de l'acquisition de connaissances, des interactions homme-machine et de la traduction automatique relayée ensuite par le traitement automatique des langues.

L'appariement des requêtes et des documents des systèmes de recherche d'information sur internet nécessite des bases lexicales et grammaticales de grande taille, permettant de résoudre les ambiguïtés potentielles (par exemple les ontologies développées par les logiciens et les sémanticiens), mais aussi des bases de données plurilingues pour pouvoir traiter des documents dans de multiples langues. Le caractère plurilingue de l'internet nécessite un développement d'outils pour une variété de langues sans commune mesure avec les travaux précédents.

Les corpus constitués par les laboratoires de linguistique servent non seulement à tester des théories ou analyses linguistiques, mais sont aussi utilisés comme bancs d'essai pour des implémentations sur ordinateur, par exemple pour les analyseurs syntaxiques. Des campagnes de comparaison d'analyseurs syntaxiques (EASY), ainsi que de méthodes d'extraction d'informations ou de fouille de texte (MUC), se font régulièrement avec la participation non seulement de chercheurs, mais aussi de partenaires industriels.

Ces corpus étiquetés ou annotés avec des informations linguistiques servent aussi comme bases de données pour l'acquisition automatique d'informations diverses (lexique, syntaxe, structure discursive) par des méthodes d'apprentissage. Ces méthodes s'avèrent très importantes pour les recherches sur le discours, où les approches symboliques sont encore beaucoup trop fragiles pour arriver à des analyses automatiques de textes en dehors de domaines restreints, et cela est encore plus vrai de l'oral. Depuis une dizaine d'années, les approches statistiques sont devenues l'axe principal des recherches en linguistique computationnelle.

Si les techniques d'analyse et de traduction automatiques de l'écrit, comme les systèmes de synthèse ou de reconnaissance de la parole, semblent bien maîtrisés, de nouveaux défis apparaissent continuellement, comme le dialogue homme-machine oral et sans interface, avec de multiples applications tels que l'enseignement, la conduite automobile ou la médecine.

D'une façon plus générale, la possibilité d'intégrer dans les systèmes de TAL la prise en compte de sources d'informations différentes (signal acoustique, structure lexicale, morpho-syntaxique, données mimo-gestuelles) ouvre le champ du traitement automatique à l'étude de la multimodalité.

## Conclusions

Comme de nombreuses sections de l'INSHS, la section 34 voit partir à la retraite un nombre croissant de chercheurs et d'ingénieurs et techniciens. En 2010 la section compte 160 chercheurs titulaires et 42 devraient quitter l'organisme d'ici 2014 (28 DR et 14 CR), soit un taux de 32%. Ces départs risquent si nous n'y prenons garde, d'obérer des pans entiers de nos disciplines, en particulier dans les domaines où l'université hésite encore à s'engager. Il faudrait recruter au moins 10 chercheurs chaque année pour maintenir la présence de chercheurs à temps plein dans les unités, en particulier de jeunes chercheurs s'engageant tôt (au niveau CR2) sur des projets nécessitant un fort investissement en temps d'enquête, de saisie et d'analyse.

Ces recrutements devraient être accompagnés d'un accroissement significatif des postes d'ITAS affectés dans les unités. Durant la période récente, les standards méthodologiques couramment admis dans la discipline ont considérablement évolué, avec des exigences accrues dans la collecte et le traitement de données attestées ou provoquées (comportementales ou de neuro-imagerie) ce qui fait que la linguistique est d'ores et déjà une discipline «appareillée» et va tendre à le devenir de plus en plus. Cette évolution implique des tâches d'ingénierie de plus en plus lourdes (mise en place de plates-formes pour l'annotation de corpus, pour le traitement des données, pour le montage d'expérimentations et pour l'élaboration d'outils TAL. Les besoins en personnels (et en équipements) sont importants et conditionnent le développement des recherches actuelles et futures susceptibles de satisfaire aux nouvelles exigences méthodologiques requises pour franchir le cap des évaluations internationales.

Une concertation plus étroite entre le CNRS et les universités permettrait de mettre en place une formation en sciences du langage solidement ancrée sur un véritable cursus commun entraînant tout autant les étudiants à la linguistique générale qu'aux enquêtes de terrain ainsi qu'aux ressources (TAL, statistiques, ...) et aux démarches expérimentales. Ces liens étroits pourraient aussi se manifester par un accroissement significatif des délégations, et par l'offre de post-docs ou bourses d'études permettant de développer des cursus multidisciplinaires.

Section à caractère interdisciplinaire, la section 34 entretient des liens étroits avec la CID 44, mais beaucoup moins avec la CID 42. Cette différence se marque tant par l'absence de représentant de la section 34 auprès de la CID 42 que par une remarquable convergence dans la sélection des doubles candidatures entre les sections 34 et 44, qui n'existe pas entre les sections 34 et 42. Dans ces conditions, la section 34 s'interroge sur le sens de sa présence au sein des sections concernées par la CID 42. Les sciences du langage sont une des composantes

essentielles des sciences humaines auxquelles elles participent par leurs analyses spécifiques, par leurs interfaces avec d'autres disciplines et par leurs applications. Dans la structuration en cours à l'échelle des régions voulue par le CNRS, dès lors que les MSHS pourraient devenir l'instance de structuration de la recherche en Sciences Humaines et Sociales, il est indispensable que les sciences du langage y trouvent leur place, non seulement les UMR qui n'y sont pas toutes intégrées mais aussi les laboratoires, les équipes ou les programmes qui se sont signalés par la qualité de leurs travaux. C'est en effet en unissant les forces des UMR et des équipes universitaires que les sciences du langage gagneront en qualité et en visibilité.

Rappelons également que deux fédérations (ILF et TUL) regroupent la majorité des laboratoires qui relèvent de la section 34. Ces deux fédérations jouent chacune dans leur domaine un rôle essentiel de cohésion, d'échange et de prospective. Leur réactivité en fait une interface performante entre les unités et la direction scientifique, comme avec le Comité national. Leurs missions doivent donc être maintenues. Toutefois, dans la mesure où les unités de recherche regroupent très souvent des chercheurs qui partagent des thématiques avec l'une ou l'autre fédération, la question de l'appartenance exclusive des laboratoires à une seule fédération, qui est actuellement la règle, mérite d'être posée, comme mérite d'être posée la question de l'ouverture des fédérations à des chercheurs qui conduisent leurs travaux dans d'autres cadres que les UMR.

Conscients d'être des acteurs sociaux de plein droit, les linguistes dérogeraient à leur mission s'ils ne menaient pas également une réflexion sur leur rôle au sein de la société et de la cité. Dans cette perspective, il n'est pas inutile de rappeler que les connaissances produites par les sciences du langage ont des applications dans divers domaines: en réponse aux besoins de politique linguistique, dans l'enseignement et l'ingénierie des langues, dans le traitement des pathologies de la voix, du langage ou de la perception auditive.

## ANNEXE 1

ANR Agence Nationale de la Recherche  
 EHESS Ecole des Hautes Etudes en Sciences Sociales  
 ENS Ecole Normale Supérieure  
 ERC European Research Council  
 GDR Groupement De Recherche  
 ILF Institut de Linguistique Française (FR2393)  
 INRIA Institut National de Recherche en Informatique et en Automatique  
 INS2I Institut des sciences informatiques et de leurs interactions  
 INSHS Institut National des Sciences Humaines et Sociales  
 JSM Journées de Sémantique et Modélisation  
 MIT Massachusetts Institute of Technology  
 PICS Projet International de Coopération Scientifique  
 SPE Semantics and Philosophy in Europe

*Rapport de conjoncture 2010*

TAL Traitement Automatique des Langues  
TUL Fédération de recherche en Typologie et Universaux  
Linguistiques (FR2559)  
UCLA University of California, Los Angeles  
UMR Unité Mixte de Recherche