
RAPPORT DE CONJONCTURE

DU COMITÉ NATIONAL DE LA RECHERCHE SCIENTIFIQUE

ÉDITION 2014

Extrait



CNRS ÉDITIONS

15, rue Malebranche – 75005 Paris

SECTION 34

SCIENCES DU LANGAGE

Extrait de la déclaration adoptée par le Comité national de la recherche scientifique réuni en session plénière extraordinaire le 11 juin 2014

La recherche est indispensable au développement des connaissances, au dynamisme économique ainsi qu'à l'entretien de l'esprit critique et démocratique. La pérennité des emplois scientifiques est indispensable à la liberté et la fécondité de la recherche. Le Comité national de la recherche scientifique rassemble tous les personnels de la recherche publique (chercheurs, enseignants-chercheurs, ingénieurs et techniciens). Ses membres, réunis en session plénière extraordinaire, demandent de toute urgence un plan pluriannuel ambitieux pour l'emploi scientifique. Ils affirment que la réduction continue de l'emploi scientifique est le résultat de choix politiques et non une conséquence de la conjoncture économique.

L'emploi scientifique est l'investissement d'avenir par excellence

Conserver en l'état le budget de l'enseignement supérieur et de la recherche revient à prolonger son déclin. Stabiliser les effectifs ne suffirait pas non plus à redynamiser la recherche : il faut envoyer un signe fort aux jeunes qui intègrent aujourd'hui l'enseignement supérieur en leur donnant les moyens et l'envie de faire de la recherche. On ne peut pas sacrifier les milliers de jeunes sans statut qui font la recherche d'aujourd'hui. Il faut de toute urgence résorber la précarité. Cela suppose la création, sur plusieurs années, de plusieurs milliers de postes supplémentaires dans le service public ainsi qu'une vraie politique d'incitation à l'emploi des docteurs dans le secteur privé, notamment industriel.

Composition de la section

Anne REBOUL (présidente de section); Nicolas QUINT (secrétaire scientifique); Anne-Marie ARGENTI; Nicholas ASHER; Mahe BEN HAMED; Sylvie BLET; Georgette DAL; Hamida DEMIRDACHE; Christine DEVELLOTTE; Sophie DUFOUR; Claire GARDENT; Judit GERVAIN; Jonathan GINZBURG; Pierre HALLE; Corinne KALFA; Brenda LACA; Sylvain LOISEAU; Cedric PATIN; Rachid RIDOUANE; Jean-Luc SCHWARTZ; Marleen VAN PETEGHEM.

Résumé

La section 34, consacrée aux sciences du langage, se répartit en 5 grands domaines : la linguistique fondamentale (phonologie ; morphologie ; syntaxe ; sémantique), les mécanismes généraux de l'usage du langage dans la communication (phonétique ; philosophie du langage ; pragmatique ; discours, texte et dialogue ; évolution du langage), la diversité des langues (typologie et diversité des langues ; linguistique diachronique et linguistique comparée ; sociolinguistique, variation et contact de langues), des approches transversales (psycholinguistique ; traitement automatique du langage naturel et linguistique informatique ; lexicographie et lexicologie), et l'histoire de la linguistique. Il a semblé utile d'ajouter une section sur la démographie des chercheurs rattachés à la 34 et sur celles des ITA affectés à des laboratoires relevant de la 34.

Introduction

Le langage, capacité propre à l'espèce humaine, est impliqué dans toutes les activités humaines et, à ce titre, toutes les sciences humaines et sociales pourraient passer, à tort, pour des sciences du langage. Il importe donc de délimiter le champ des sciences du langage afin que toute activité scientifique dans laquelle le langage intervient d'une façon ou d'une autre ne s'y retrouve pas.

Les sciences du langage s'intéressent au langage comme capacité humaine universelle, intimement liée à des processus cognitifs d'une part et à des processus sociaux de l'autre. Ce sont les caractères structurels du langage ainsi que les processus généraux qui en régissent l'emploi, notamment dans la communication, qui constituent le cœur des sciences du langage. L'universalité du langage ne contredit cependant en rien la diversité des langues naturelles qui constitue aussi un sujet de recherche important au sein des sciences du langage. On distinguera encore des approches transversales, comme la psycholinguistique, le traitement automatique des lan-

gues, ou la lexicographie, qui peuvent toucher n'importe lequel des domaines évoqués précédemment et se distinguent de la linguistique fondamentale ou de l'étude de la diversité linguistique en ce qu'ils utilisent des méthodologies différentes, dont certaines peuvent déboucher sur des applications concrètes, dans l'industrie, la santé, ou l'éducation. Enfin, un petit nombre de chercheurs se consacrent à l'histoire de la discipline. Pour finir, on donnera un aperçu général de la démographie des chercheurs et des ITA CNRS affiliés à la section.

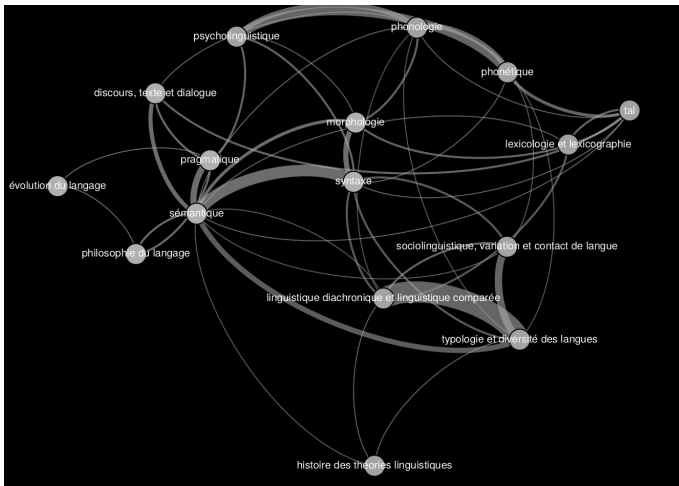
On remarquera à cet égard que, pour chaque section ou sous-section, le pourcentage de chercheurs participant au domaine concerné est indiqué dans le titre. Ceci ne signifie cependant pas que les domaines soient discontinus et la plupart des chercheurs contribuent à plusieurs domaines ou sous-domaines. Cette plurivalence est indiquée dans le schéma ci-après (on trouvera dans la Section VI des informations sur les données sources) où la communauté de couleurs entre plusieurs sous-disciplines marque le fait que des chercheurs partagent leur activité entre ces différents domaines.

I. Linguistique fondamentale (~ 31 %)

La linguistique fondamentale s'intéresse aux structures universelles du langage, à leurs caractéristiques formelles et à la contribution qu'elles font à la signification des énoncés. À ce titre, les disciplines qui constituent cette thématique contribuent à l'investigation des caractères fondamentaux de cette capacité spécifique à l'espèce humaine et constituent le cœur des sciences du langage.

A. Phonologie (~ 6 %)

La thématique *Phonologie* se retrouve dans de nombreux laboratoires : LPL (UMR7309), SFL



La structure des disciplines et sous-disciplines au sein de la section 34. À partir des mots-clés fournis par les chercheurs pour qualifier leurs propres thématiques de recherches, une nomenclature en 16 sous-disciplines a été construite, puis un graphe de co-occurrence entre catégories a été établi (l'épaisseur des liens est proportionnelle à la fréquence de co-occurrence des catégories chez un même chercheur).

(UMR7023), CLLE (UMR5263), CRLAO (UMR8563), LLF (UMR7110), MoDyCo (UMR7114), LPP (UMR7018), LSCP (UMR8554), BCL (UMR 7320), DDL (UMR5596), STL (UMR8163), GIPSA (UMR5216).

La thématique *Phonologie* est consacrée à l'identification et à l'étude des sons pertinents du langage, ainsi que des principes qui régissent leurs distributions et fonctions. Par sa dimension symbolique, elle entretient naturellement des liens avec la morphologie (plus particulièrement dans le cadre du modèle CVCV), la syntaxe et la sémantique mais, parce qu'elle consiste en l'interprétation de données phonétiques, c'est avec la *Phonétique* que l'on observe une forte proximité, tout particulièrement depuis l'émergence du courant de la *Phonologie de Laboratoire*, auquel peuvent être associés la moitié des chercheurs de la thématique. Ce cadre d'analyse appuie les analyses phonologiques sur des analyses phonétiques précises relevées via l'expérimentation, la modélisation et, de plus en plus, par l'examen de corpus (de grande taille). La démarche expérimentale en phonologie est aussi fortement visible dans les travaux employant les méthodes de la psychologie expérimentale (voir § IV.A). 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 abstrait de la parole humaine et son aspect physique et traitent principalement de questions liées à la nature des représentations phonologiques chez l'adulte et l'enfant. À l'inverse, plusieurs chercheurs tendent à dissocier forme et substance dans leurs analyses et modélisations, qu'ils inscrivent dans des cadres formels, génératifs ou post-génératifs le plus souvent. Au niveau théorique, c'est l'approche partagée du modèle CVCV qui fédère la plupart de ces travaux, qui jouissent d'une visibilité internationale indéniable. Certains chercheurs travaillent sur des langues non indo-européennes, et tout particulièrement sur les langues afro-asiatiques (berbère) et les langues sino-tibétaines (cf. § III).

B. Morphologie (~ 4%)

La thématique *Morphologie* est représentée au LLF (UMR7110), dans l'équipe Equipe de Recherche en Syntaxe et Sémantique du CLLE (UMR 5263), au SFL (UMR 7023) et au STL (UMR 8163).

La morphologie vise à étudier les corrélations régulières qu'entretiennent les lexèmes avec leurs mots-formes (on parle alors de morphologie flexionnelle) ou les lexèmes entre

eux quand les uns présentent un degré de complexité structurelle et sémantique supérieur aux autres (il s'agit alors de morphologie constructionnelle).

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 lexèmes et de leurs mots-formes, les recherches s'inscrivent également dans le champ de la morphophonologie, de la sémantique, de la syntaxe, de la pragmatique ou de la psycholinguistique. La morphologie, qui se prête par ailleurs bien à la modélisation (largement utilisée dans cette thématique), donne également lieu à des travaux relevant du traitement automatique des langues et de la linguistique computationnelle.

Toutes ces orientations se retrouvent dans l'éventail des recherches actuellement menées en France, dans des cadres théoriques variés, qui vont de la morphologie lexématique à la morphologie dite distribuée. Les méthodologies pour lesquelles ces travaux s'appuient vont de l'étude de corpus à la linguistique expérimentale.

Les recherches en morphologie, outre le français ou les langues romanes, sont aussi conduites sur des langues non indo-européennes (notamment sur des langues afro-asiatiques, amérindiennes, sino-tibétaines et sur le basque).

C. Syntaxe (~ 7%)

On notera une très forte concentration des recherches en syntaxe dans des laboratoires parisiens : notamment SFL (UMR7023) et LLF (UMR7110) pour la syntaxe formelle, le CRLAO (UMR8563) pour la syntaxe des langues d'Asie Orientale, ou le LLACAN (UMR8135) pour la syntaxe des langues africaines. IKER (UMR5478) à Bayonne et le LLING (EA3827) à Nantes sont les seules équipes de province.

La syntaxe étudie la façon dont des unités de signification (morphèmes, mots) se combinent entre elles pour générer des unités de signifi-

cation plus grandes (syntagmes, propositions, phrases). Elle cherche à déterminer les combinaisons possibles (et impossibles) d'unités et à découvrir les principes qui président à leur combinaison, et ce faisant à éclairer les mécanismes de leur interprétation. De par le principe de compositionnalité du sens (qui dit que le sens d'une expression complexe est une fonction du sens des unités qui la composent et de leur mode de combinaison), une théorie adéquate de (la compositionnalité de) la signification doit être articulée sur une théorie de la syntaxe.

La question de l'articulation entre syntaxe et sémantique est au cœur des controverses qui opposent les grands courants de syntaxe formelle. On peut (grossièrement) distinguer deux grands paradigmes théoriques. D'un côté, les approches lexicalistes et monostratales (Grammaires Catégorielles, *Lexical Functional Grammar* (LFG), *Head-driven Phrase Structure Grammar* (HPSG)) où le calcul sémantique s'opère sur des structures de « surface » (transparentes), et qui font appel à une sémantique puissante (enrichie) pour rendre compte de la (non-)correspondance entre syntaxe et sémantique. De l'autre, les modèles comme la Grammaire Générative où pour rendre compte de la (non-) correspondance entre structures syntaxiques et sémantiques, on fait appel à une composante syntaxique puissante (abstraite) qui génère des « Formes Logiques transparentes » pour le calcul sémantique.

Les recherches en syntaxe sont bien diversifiées, tant du point de vue des paradigmes scientifiques et méthodologiques représentés, qu'au niveau des aires culturelles couvertes. Ainsi, en ce qui concerne les cadres théoriques adoptés, le CNRS compte de très bons chercheurs en grammaire générative, en HPSG, ou en grammaire des contraintes, menant des recherches de pointe en syntaxe à l'interface avec d'autres niveaux d'analyse, notamment la sémantique et la morphologie, et plus rarement la phonologie ou la prosodie. La section compte aussi plusieurs membres qui combinent la syntaxe avec le TAL, la psycholinguistique, la typologie ou la sociolinguistique. Ces recherches se distinguent également par le

recours à une méthodologie expérimentale pour collecter des données de compréhension ou de production et/ou mettre à l'épreuve des hypothèses théoriques, ainsi que par le recours à des méthodes d'analyse quantitative, ou par l'usage de corpus.

On soulignera, enfin, que des recrutements récents ont ouvert la porte à des paradigmes de recherche en syntaxe nouveaux au CNRS, et offrant des questionnements nouveaux, ou des perspectives renouvelées sur certaines questions classiques de la syntaxe : syntaxe et sémantique des langues signées, neuro-syntaxe (corrélats neuronaux des structures hiérarchiques), et biolinguistique explorant les bases biologiques du langage.

Pour ce qui est des langues étudiées, les recherches portent sur une belle palette de langues : langues romanes, germaniques ou slaves, mais également des langues typologiquement plus éloignées, notamment le basque, les langues d'Asie orientale, langues celtiques, ainsi que les créoles, les langues mandéées, les langues tchadiques, ou la langue des signes.

Quoique le vivier de chercheurs en syntaxe au CNRS soit réduit, la qualité des chercheurs dans ce domaine est incontestable, et leurs recherches ont une audience internationale importante.

D. Sémantique (~ 14%)

La sémantique est pratiquée un peu partout sur le territoire, mais on citera notamment pour les approches formelles vériconditionnelles dites plus traditionnelles, Paris (avec l'IJN, UMR8129 ; SFL, UMR7023 ; LLF, UMR 7110 ; LATTICE, UMR8094 ; CRLAO, UMR8563), Bayonne (IKER, UMR5478) ou Nantes (LLING, EA3827). Pour la sémantique dynamique on note Toulouse (IRIT, UMR5505 ; CLLE, UMR5263), Nancy (LORIA, UMR7503) et Bordeaux (LABRI, UMR5800), pour les approches formelles alternatives Paris (LLF) et Toulouse (IRIT), pour la sémantique computationnelle,

le LORIA, l'IRIT et pour des approches descriptives, la région parisienne (LACITO, UMR7107 ; LLACAN, UMR8135).

La sémantique a pour sujet le sens ou contenu d'un mot ou d'une construction dans une langue donnée. Elle traite aussi la composition de sens pour donner un contenu aux phrases, voire à des textes entiers. Son champ correspond essentiellement à 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, de l'ellipse et des anaphores, et de la relation entre la sémantique lexicale et la structure argumentale 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, l'implicature, ou la dimension émotive ou évaluative. La sémantique suit une méthodologie largement formelle, bien que certaines approches restent descriptivistes.

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, comme celui de la logique des prédicats (langage du premier ordre) ou de logiques d'ordre supérieur. Dans cette modélisation, la valeur d'une phrase à un indice d'évaluation (qui peut inclure un moment temporel, une situation possible, un contexte déictique, un contexte discursif, *inter alia*) est une valeur de vérité, calculée récursivement à partir des dénominations assignées aux composantes de la phrase. C'est pourquoi cette approche sémantique s'appelle la sémantique vériconditionnelle.

La sémantique est actuellement dans une période de développement et de changement profonds vis à vis du paradigme dominant de la sémantique vériconditionnelle. Elle est en effet

en train d'assimiler des méthodes et des concepts issus de disciplines comme l'informatique et la psychologie cognitive. On note ainsi un essor important de la sémantique computationnelle qui conduit, par exemple, à l'exploitation de théories de types riches, empruntées à la sémantique des langages de programmation, dans le cadre de la sémantique des langues naturelles, en France et en Europe. Cet axe de recherche a aussi renouvelé, en raison de l'étroite correspondance entre types et preuves, un intérêt pour la théorie de la preuve comme cadre théorique pour la modélisation du sens.

Un autre développement important est l'émergence de théories probabilistes de la sémantique, 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.

On observe aussi l'importance croissante du développement de modèles sémantiques issus de l'approche distributionnelle, qui essaient de caractériser le sens d'un mot et même de constructions plus complexes par des méthodes issues de l'algèbre linéaire. Le défi majeur est de produire une théorie de la composition du sens à partir de représentations issues de l'algèbre linéaire, ce qui constitue un domaine de recherche très actif dans la communauté mondiale.

On constate dans le paradigme formel des ouvertures importantes vers 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) d'approches empiriques nouvelles, et ceci dans les deux domaines que constituent le travail sur les grands corpus et les expériences psycho-, voire neuro-linguistiques ; (iii) 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 déixis,

d'attitudes propositionnelles et d'ellipse ou d'anaphore. On notera aussi une poussée au delà des textes vers une analyse sémantique et pragmatique du dialogue (cf. § II.D). Bien que le français, les langues romanes et l'anglais fournissent le gros des données analysées dans les approches formelles, il y a également une production soutenue d'analyses de qualité sur l'interface syntaxe-sémantique dans des langues typologiquement variées (basque, langues d'Asie Orientale, etc.). En revanche, le travail sur des langues peu étudiées ou en danger, qui a beaucoup contribué à faire découvrir ou reformuler certaines questions de recherche, n'en est qu'à ses débuts en France et reste souvent descriptif.

Le domaine de la sémantique en France comprend donc aussi des recherches en dehors du cadre de la sémantique formelle au sens large, à savoir de la sémantique descriptive qui s'inscrit dans des cadres généralement cognitivistes ou fonctionnalistes, mais qui reste néanmoins théoriquement assez hétérogène.

La communauté française en sémantique formelle s'inscrit avec succès dans le débat scientifique international.

II. Mécanismes généraux de l'usage du langage dans la communication (~ 24 %)

Le langage est au centre de la communication humaine et, à ce titre, l'étude de son usage pour cette fonction complète la linguistique fondamentale. Les mécanismes généraux de l'usage du langage dans la communication s'articulent à des processus cognitifs liés à la perception et à la capacité d'intégration d'informations linguistiques à des informations non-linguistiques, notamment à des informations contextuelles.

A. Phonétique (~ 13 %)

La thématique *Phonétique* est représentée dans trois laboratoires principaux, GIPSA-Lab (UMR5216), LPL (UMR7309) et LPP Paris 3 (UMR7018), ainsi qu'au LIMSI (UPR3251) et à l'Institut de Phonétique de Strasbourg (EA 3403). 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 principalement la description et la modélisation de la perception et, le plus souvent, de la production de parole du point de vue acoustique et/ou articuloire (e.g. 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 la phonétique et la phonologie de laboratoire ou la psycholinguistique (notamment sur l'acquisition et le bilinguisme). D'autres axes de recherche concernent la description des systèmes sonores des langues, en lien avec la phonologie et la typologie, ou bien du point de vue de leur évolution (e.g. variétés régionales des langues et dialectes romans); l'analyse de la prosodie des langues; les pathologies de la parole.

La plupart des recherches en phonétique sont cependant de type expérimental (production: analyses acoustiques et/ou de l'articulation; perception: tests comportementaux, électrophysiologie, imagerie cérébrale). Le recueil et l'analyse de corpus (de contrôlé à spontané) est largement représenté, avec, pour l'acquisition de la parole, des corpus de parole enfantine ou de parole adressée aux enfants. La modélisation (acoustique et/ou articuloire) est également présente dans les études de production. Quelques chercheurs sont spécialisés dans des langues particulières (e.g., langues afro-asiatiques, langues de l'Asie du sud-est), mais la majorité sont généralistes ou comparatistes.

B. Philosophie du langage (~ 1 %)

La philosophie du langage se concentre principalement à l'IJN (UMR8129) et au laboratoire IHPST (UMR8590) à Paris, mais aussi à STL (UMR8163) à Lille et à L2C2 (UMR5304).

La philosophie du langage, qui se développe dans les années 1950-60 à Oxford, a donné naissance à la pragmatique, avec deux courants principaux à l'époque, la théorie des actes de langage et l'émergence de théories portant sur la communication implicite (présuppositions et implicatures). La philosophie du langage actuelle se rattache maintenant au moins autant à la philosophie analytique du début du xx^e siècle, étendant son champ à la sémantique et à la logique, tout en conservant une forte préoccupation pour des thématiques pragmatiques, principalement liées à la communication implicite. L'intérêt pour la communication implicite a contribué de façon cruciale au débat entre les approches sémantiques minimalistes et le contextualisme, débat qui impacte directement la localisation de la frontière entre sémantique et pragmatique dans l'interprétation des énoncés. Par ailleurs, la philosophie du langage a développé un volet important sur les rapports entre langage et pensée, et on pensera ici plus précisément aux travaux sur la référence et les dossiers mentaux, à l'implication de la perspective dans la production des énoncés, notamment des énoncés méta-linguistiques ou à de nouvelles approches du vague articulant modèles psychologiques et modèles logiques. On mentionnera enfin l'établissement de liens entre langage et ontologie, et le développement d'investigations philosophiques reposant sur des modèles linguistiques, notamment sémantiques.

Les travaux en philosophie du langage ont un impact important (et réciproque) sur d'autres thématiques en sciences du langage, et on citera ici bien évidemment la sémantique et la pragmatique.

L'impact international de ces travaux, ainsi que leur qualité, est particulièrement remarquable, vu le petit nombre de chercheurs et d'enseignants-chercheurs de la discipline.

C. Pragmatique (~ 3 %)

En France, la pragmatique est principalement représentée à Paris (LSCP UMR8554; IJN UMR8129), Lyon (L2C2 UMR5304) et Toulouse (IRIT UMR5505; CLEE UMR5263).

Par contraste avec la sémantique, qui se concentre sur la façon dont le sens d'une phrase est obtenu par la composition du sens de ses composants articulée par les structures syntaxiques (cf. § *Syntaxe*), la pragmatique s'intéresse à des composants de l'interprétation qui échappent en partie aux processus compositionnels et dépendent de différents éléments contextuels, lesquels couvrent la situation de communication, l'interprétation des énoncés précédents, ainsi que les connaissances encyclopédiques. La pragmatique n'est pas liée à l'analyse de discours, bien qu'elle puisse s'intéresser à certains aspects du discours ou du dialogue (cf. § II.D). Elle s'inscrit en partie dans le cadre du contextualisme (cf. § II.B) et s'intéresse en particulier à la communication implicite (présuppositions et implicatures conventionnelles et conversationnelles), et à la façon dont les interlocuteurs récupèrent les contenus implicites. La communication implicite est un domaine central de l'interface sémantique-pragmatique et les contributions respectives de la combinatorialité syntactico-sémantique et du contextualisme dans l'interprétation font l'objet d'un débat important à l'heure actuelle. On peut distinguer deux approches principales des problèmes pragmatiques, compatibles et souvent pratiquées par les mêmes chercheurs, l'approche formelle, principalement développée dans les laboratoires parisiens et toulousains, et l'approche expérimentale, initiée à Lyon dans le cadre d'un projet ESF (European Science Foundation) X-Prag, puis adoptée également dans les laboratoires parisiens. L'approche expérimentale

utilise les méthodes, principalement comportementales, mais incluant aussi de l'électroencéphalographie, de la psychologie expérimentale (cf. § IV.A). Elle a largement permis d'écarter certaines approches théoriques au profit d'autres et évolue avec les modifications théoriques qu'elle a largement contribué à produire. On pensera notamment ici à la dernière version de la théorie des alternatives proposée par Chierchia. Son retentissement au niveau européen et international est important et certains pays ont développé des appels à projet qui lui sont spécifiquement dédiés (e.g., le programme X-prag.de, Deutsche Forschungsgemeinschaft).

L'approche formelle, qui intègre de plus en plus souvent une partie expérimentale, est également très bien insérée dans le débat international, comme en témoigne la grande qualité des publications produites.

Il s'agit donc là d'un domaine qui, malgré le petit nombre de chercheurs impliqués, est d'une grande qualité et a un impact international majeur.

D. Discours, texte et dialogue (~ 6 %)

La rubrique Discours, texte et dialogue concerne les approches linguistiques portant sur la structure et le contenu de productions linguistiques plus longues qu'une phrase, et incluant souvent la multidimensionalité et la multimodalité. Une hypothèse commune aux travaux dans ce domaine est l'idée que la structure de telles productions linguistiques n'est pas seulement une suite de phrases ou d'énoncés, et que leur contenu n'est pas seulement une conjonction ou une intersection des valeurs sémantiques de ces phrases. On trouve en France plusieurs approches dans ce domaine : notamment à Toulouse (IRIT, UMR5505), Paris (LLF, UMR7110, LATTICE, UMR8094 et ALPAGE, UMRI-001), Caen (GREYC, UMR6072), Nancy (LORIA, UMR7503) et Aix-Marseille (LPL, UMR7309)

pour les approches formelles et computationnelles ; à Lyon (ICAR, UMR5191), Montpellier (Praxiling, UMR5267), Aix-Marseille (LPL, UMR7309), Villejuif (SEDYL, UMR8202) pour les approches interactionnelles.

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, qu'en dehors des sciences du langage, en lien avec les sciences cognitives, la psychologie et les sciences sociales pour la modélisation des paramètres sociaux-cognitifs impliqués dans l'acte communicationnel.

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 oeuvre 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 et 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 qui sont analysées sont situées dans des contextes sociaux spécifiques (éducatifs, commerciaux, médicaux..) ou encore médiées par la technologie. L'un des enjeux forts de ces dernières recherches consiste à 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.

On observe des efforts croissants afin de fournir des analyses formelles de l'interaction dialogique, pour capter en une manière précise les aperçus empiriques de la linguistique

interactionnelle et de la psychologie cognitive, mais aussi en collaboration avec de la recherche en TAL pour sous-tendre les systèmes de dialogue parlés (e.g. le système SIRI d'Apple). De même, depuis plusieurs années, des efforts se sont concentrés sur une interaction productive entre les études formelles de la structure et du contenu du discours, comme dans la SDRT, RST ou DLTAG, et les approches computationnelles. Les conférences majeures en TAL comme ACL, EMNLP, et EACL (cf. Annexe), ont depuis quelques années des sections dédiées à l'extraction automatique de la structure discursive d'un texte et des modes d'interaction dialogique, en exploitant des méthodes statistiques d'apprentissage automatique ou hybrides (approches utilisant à la fois les principes symboliques des études formelles sur le discours/dialogue et des méthodes statistiques).

E. Évolution du langage **(~ 1 % des effectifs de la section)**

L'évolution du langage (et non pas l'évolution des langues ou l'acquisition linguistique) est un problème par nature profondément interdisciplinaire, à la frontière de la biologie, de la paléo-anthropologie, des sciences du langage et de la psychologie comparée. Après un siècle d'indifférence, la question est revenue sur le devant de la scène scientifique internationale dans les années 1990 et fait maintenant l'objet de colloques dédiés (e.g. Evolang) et de nombreuses publications internationales. La France contribue à ce développement et les recherches se poursuivent principalement à Lyon (L2C2 UMR5304, DDL UMR5596), Grenoble (GIPSA, UMR5216) ainsi qu'à Paris (Paris-tech, Telecom) et Bordeaux (INRIA) en relation avec l'informatique et la robotique cognitive.

On peut distinguer deux grands types d'approche 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, e.g., la création de conventions lexicales ou étudiant sur cette

base les 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, mais pas toujours, basées sur la théorie des jeux. Elles sont peu représentées en France. Il y a d'autre part des approches de nature plus théorique et interdisciplinaire, cherchant sur la base des nombreux travaux empiriques existants, à dégager les spécificités structurelles du langage ainsi que ce qui est propre au langage utilisé dans la communication humaine par rapport aux systèmes de communication animaux (psychologie comparée). L'objectif est, sur la base de ces spécificités et des capacités cognitives (et culturelles) humaines qui leur semblent liées, de dégager des scénarios pour l'évolution du langage. Ce courant se base sur la psychologie cognitive, les neurosciences, la paléo-anthropologie, l'anthropologie, la simulation informatique, et les sciences du langage. C'est là que s'inscrivent les travaux français dans leur majorité.

Ces travaux ont une place honorable dans la recherche internationale, mais qui gagnerait à augmenter au vu de l'importance de cette thématique profondément interdisciplinaire dans la science actuelle. Pour ce faire, l'enjeu du développement des relations interdisciplinaires entre sciences du langage, neuro-psychologie cognitive et sociale, primatologie comparée et sciences de l'information est crucial.

III. Universalité du langage, diversité des langues (~ 26%)

On compte environ 6500 langues différentes, qui présentent une grande diversité, tant formelle que systémique.

Étant donné que de nombreuses langues du monde restent peu ou pas décrites, que beaucoup d'entre elles, essentielles pour notre compréhension de la diversité linguistique, sont en train de disparaître (la moitié des langues aujourd'hui pratiquées sont menacées à l'horizon 2100) et que, même pour les langues les plus décrites, seuls certains lectures (par exemple la langue écrite ou des variétés de prestige pratiquées par les groupes socio-culturels dominants) sont connus en profondeur, la production, à partir de données recueillies sur le terrain (ou dans les documents écrits disponibles) de descriptions englobantes et fidèles (grammaires, dictionnaires, textes transcrits et annotés, études sociolinguistiques, atlas dialectologiques), d'ouvrages comparatifs et de reconstructions est une activité fondamentale et nécessaire au progrès de nos connaissances sur les langues dans leur diversité et au développement des Sciences du Langage.

Le CNRS a une longue tradition de recherche dans ce domaine (langues africaines, amérindiennes, sino-tibétaines et océaniques) via des laboratoires principalement en région parisienne, comme le CRLAO (UMR8503), le LACITO (UMR7107), le LLACAN (UMR8135), le LPP (UMR7018) ou le SEDYL (UMR8202), mais aussi à Lyon (DDL, UMR5596), à Bayonne (IKER, UMR5478), à Aix en Provence (IREMAM, UMR7310).

A. Typologie et diversité des langues (~ 12%)

La typologie linguistique est l'étude des régularités attestées à travers les langues du monde, permettant de définir des universaux linguistiques, c'est-à-dire des propriétés véritablement partagées par toutes les langues du monde. À partir de ces universaux, la typologie linguistique vise à identifier plus généralement des types de langues dont la mise au jour permet de progresser dans la caractérisation des processus cognitifs sous-jacents à l'élabo-

ration et au fonctionnement des systèmes linguistiques.

La typologie des langues peut s'appuyer sur tous les niveaux de description : phonétique, phonologique, morphologique, syntaxique, sémantique, etc. De nouveaux domaines sont désormais investis par cette discipline. Ainsi, la typologie linguistique s'est notamment intéressée récemment à la comparaison des langues en ce qui concerne leur mode d'expression de certaines catégories cognitivement centrales, comme l'espace et le temps. La méthodologie des cartes sémantiques, qui consiste à observer quelles valeurs sémantiques sont « codées ensemble » (c'est-à-dire au moyen d'un même signifiant : ainsi, l'instrumental et le comitatif sont souvent exprimés au moyen du même morphème – *avec* en français) dans les langues, est un autre domaine d'élargissement de la typologie linguistique. Ces deux exemples illustrent l'intérêt grandissant en typologie linguistique pour les catégories sémantiques (lexicales et grammaticales). Notons aussi la prise en compte croissante de la prosodie dans les études typologiques (caractérisation de l'intonation des séquences interrogatives et de celles marquant le topique ou le focus à partir de corpus oraux).

La recherche des universaux langagiers que propose la typologie s'appuie sur un effort de description de la diversité linguistique et a significativement contribué à renouveler ce champ d'activité. Ainsi, le besoin des typologues d'avoir des données à la fois comparables et fiables a été l'un des moteurs de la rationalisation et de la standardisation de certains aspects de la description – comme l'adoption de systèmes d'interlinéarisation communs, inspirés des *Leipzig Glossing Rules*.

Parmi ces renouvellements des pratiques, on peut également mentionner le développement de la linguistique documentaire, qui met l'accent sur le recueil et l'archivage pérenne de données riches, permettant de préserver un témoignage des langues menacées et permettant d'éventuelles descriptions futures.

L'effort de description et de documentation des langues du monde a permis récemment la

constitution de vastes bases de données internationales recensant les propriétés d'un grand nombre de langues du monde (cf. le *World Atlas of Language Structures*). Le domaine de la typologie quantitative s'appuie sur ces vastes répertoires et s'inscrit dans un mouvement général de développement des méthodes quantitatives en linguistique. La typologie linguistique s'appuie également aujourd'hui sur des corpus textuels, en particulier grâce à la disponibilité de certains textes traduits dans un grand nombre de langues.

B. Linguistique diachronique et linguistique comparée (~ 10 %)

L'étude de l'évolution des langues dans le temps et la reconstruction des protolangues dont les langues d'une même famille sont issues est l'un des domaines les plus anciens des Sciences du langage, puisque l'hypothèse d'une langue ancestrale indo-européenne (dont sont dérivées entre autres le grec, le latin et le sanskrit) a été formulée dès 1786 par Sir William Jones. La linguistique diachronique et comparée est donc un domaine cumulatif et ancien mais il reste producteur de renouvellements théoriques et méthodologiques.

L'investissement du CNRS est conséquent dans le domaine de la comparaison des langues entre elles et de leurs évolutions diachroniques. Les chercheurs de la section 34 jouent un rôle moteur au niveau mondial dans la classification et la reconstruction des stades anciens de divers groupes génétiques : langues austronésiennes, sino-tibétaines, Niger-Congo, méso-américaines... Des travaux sont aussi consacrés à la diachronie de familles de langues indo-européennes représentées sur le territoire français (roman et celtique notamment).

La théorie de la grammaticalisation continue à jouer un rôle important et structurant dans le domaine de la linguistique diachronique. Elle permet de décrire de façon plus

rigoureuse les changements grammaticaux et sémantiques en observant des régularités et en les rapportant à des propriétés cognitives.

De nombreux renouvellements sont issus de la prise en compte des acquis de la typologie linguistique, qui permet d'identifier des régularités dans les processus de changement. Un autre renouvellement important est fourni par la *sociolinguistique historique*, qui s'attache à prendre en compte en diachronie les acquis de la sociolinguistique. Cette approche a conduit à l'exploitation de matériaux jusqu'alors peu étudiés – textes de locuteurs « peu lettrés » – mais fournissant des témoignages précieux sur la langue populaire de l'époque considérée.

Plusieurs apports méthodologiques ont renouvelé le champ des études comparatives et diachroniques. Il s'agit notamment de la numérisation à grande échelle du patrimoine écrit des langues documentées à date ancienne, qui a permis de réviser les dates de premières attestations de nombreuses formes linguistiques. Les méthodes statistiques jouent également un rôle de plus en plus important. Appliquées aux corpus numérisés, elles permettent d'étudier le caractère graduel du changement. Appliquées à des bases de données notant les innovations partagées entre langues apparentées, elles permettent de faire des hypothèses sur les états de langues anciens et de modéliser les échanges et les migrations ayant impliqué des communautés voisines.

Le CNRS dispose également d'un pôle d'expertise dans le domaine des langues anciennes (attestées essentiellement dans des documents écrits). Dans ce domaine, les travaux des membres du CNRS ont en particulier contribué à faire progresser de façon significative le déchiffrement de plusieurs de ces langues (chinois archaïque, méroïtique, tangoute) et à mettre à profit les apports de la linguistique contemporaine (typologie et linguistique générale) pour stimuler leurs champs de recherche respectifs.

Les spécialistes de ce domaine sont peu nombreux (un ou deux au maximum par langue) et généralement bien intégrés dans des réseaux de recherche internationaux (en

particulier européens), seuls à même de permettre une bonne diffusion de leurs recherches.

C. Sociolinguistique, variation et contact de langues (~ 4%)

La sociolinguistique est la branche de la linguistique qui souligne la dimension sociale de l'usage du langage, révélant l'existence d'un important degré de variation interne au système. Ces variations (phonologiques, syntaxiques, prosodiques, etc.) obéissent à différents types de structuration : on peut ainsi identifier au sein d'une même langue des lectes (ou variétés) distingués par leur caractéristiques régionales, sociales ou contextuelles. L'étude de la variation s'est renforcée avec la disponibilité de corpus permettant de synthétiser les régularités de la parole à grande échelle. On a ainsi mis en évidence l'importance des « genres textuels », associés aux différentes sphères d'activité, dont on a montré qu'ils correspondaient parfois à des sous-systèmes d'une langue.

Les approches sociolinguistique et variationnelle (développées au CNRS à ICAR UMR5191 ; PRAXILING UMR5267 ; SEDYL UMR8202 ; BCL UMR7320 ; LIMSI UPR3251) ont largement contribué à remettre en cause le mythe de l'homogénéité linguistique.

La diversité des langues se retrouve donc aussi à l'intérieur de chaque langue et la prise en compte systématique de cette variation est fondamentale pour une compréhension en profondeur des mécanismes du langage humain. La sociolinguistique contribue ainsi à la mise au point de méthodes d'analyses plus rigoureuses des données recueillies et traitées par les linguistes descripteurs ou typologistes. Elle a permis d'améliorer les techniques de collecte (en particulier au niveau de la documentation linguistique) et d'établir des ponts interdisciplinaires avec des disciplines connexes telles que la sociologie, l'ethnologie et l'anthropologie. La sociolinguistique a aussi des applications sociales directes, en permettant de mieux

prendre en compte l'inégalité des locuteurs (donc des citoyens) face à la norme (la variété prestigieuse prescrite dans le cadre scolaire ou administratif). Par ailleurs, le caractère multifactoriel (évoqué ci-dessus) des variations étudiées par les sociolinguistes a abouti à l'établissement de corpus spécifiques et à des développements stimulants dans le domaine de la modélisation de la variation.

Le contact de langues est un domaine particulièrement crucial pour les sociolinguistes, permettant de mieux comprendre et modéliser les influences qu'exercent les uns sur les autres les divers codes dont dispose une communauté ou un individu (notion de répertoire) et l'émergence de nouvelles variétés langagières (créoles, pidgins, variétés de contact), phénomène qui renvoie à la question fondamentale de la genèse et de l'origine du langage humain.

IV. Approches transversales (~ 23 %)

Les approches transversales se distinguent de la linguistique fondamentale et de l'investigation de l'usage du langage dans la communication par leur méthodologie davantage que par leurs objets d'étude. Elles peuvent donc concerner tous les domaines indiqués plus haut et ont pour particularité de pouvoir, plus directement que les trois thématiques précédentes, se prêter davantage à des résultats applicatifs.

A. Psycholinguistique (~ 14 %)

La thématique *Psycholinguistique* regroupe une trentaine de chercheurs, dans plusieurs laboratoires : LPL (UMR7309), SFL (UMR7023), LPP Paris 3 (UMR7018), LPP Paris 5 (UMR8242), LSCP (UMR8554), GIPSA (UMP5216), L2C2 (UMR5304), DDL (UMR5596), MODYCO

(UMR7114). Cette thématique vise à décrire et à analyser les mécanismes qui permettent de produire et comprendre le langage au sens large. Elle recouvre principalement des recherches sur le langage parlé, 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, et ce très majoritairement en perception plutôt qu'en production de parole. Une partie des recherches concerne aussi le traitement du langage écrit, pour l'influence de l'orthographe sur la perception phonétique ou l'accès au lexique, pour les processus de médiation phonologique, ou encore pour l'analyse morphologique. Les psycholinguistes de la section 34 s'intéressent à l'adulte monolingue, mais aussi de plus en plus aux multilingues, avec quelques chercheurs spécialisés dans les études interlangues, ainsi que sur les langues signées et le bilinguisme bimodal. Un nombre croissant de chercheurs (8/29 en 2014) sont spécialisés dans les processus d'acquisition du langage parlé (sons élémentaires, lexique, morphosyntaxe, etc.) chez les très jeunes enfants. Les recherches sur les pathologies (dysarthries, dysphasies, dyslexies) sont une autre tendance forte. L'éventail des méthodes expérimentales de la psycholinguistique est bien représenté dans la section 34 (méthodes comportementales et électrophysiologiques, imagerie cérébrale, enregistrements articulatoires, modélisations), avec une importance notable des travaux sur corpus parlés, spontanés ou non, ou encore sur corpus écrits. Les recrutements depuis 2010 montrent que (1) l'expérimentation occupe une place de plus en plus importante, (2) huit des psycholinguistes de la section travaillent dans un laboratoire de psychologie et/ou de neurosciences cognitives et les autres dans des laboratoires orientés vers la phonétique et la phonologie (au LPL), (3) beaucoup d'entre eux travaillent sur l'acquisition.

La psycholinguistique française connaît un succès international indéniable, comme le montre la qualité des publications du domaine.

B. Traitement automatique du langage naturel et linguistique informatique (~ 6%)

Le Traitement Automatique des Langues (TAL) et la linguistique informatique (LI) traitent de la modélisation informatique des langues humaines aussi dites langues naturelles. À ce titre, on trouve des chercheurs en TALN sur deux sections du CNRS, la 34 et la 07 (Sciences de l'information). Ces chercheurs se répartissent assez uniformément sur le territoire français, de la région parisienne (LIMSI, UPR3251 ; LATTICE, UMR8094) à Nancy (LORIA, UMR7503), Toulouse (IRIT, UMR5505 ; CLLE UMR5263), Aix-Marseille (LIF, UMR7279 ; LPL, UMR7309), Avignon (LIA, EA4128) et Nantes (LINA, UMR6241).

Ces deux sous-disciplines des sciences du langage (TAL et LI) couvrent un spectre allant de l'analyse et de la modélisation des propriétés linguistiques et computationnelles des langues humaines au développement de modules et de systèmes informatiques traitant de tâches applicatives spécifiques. Au pôle théorique de ce spectre, on trouve le développement d'analyseurs syntaxiques permettant de prédire la structure en constituants d'une phrase, et à l'autre des applications telles que *Google Translate* ou le système de dialogue Homme-Machine de GDF.

Plus généralement, les grands axes de recherche de la linguistique informatique portent sur la modélisation des processus et des représentations des différents niveaux linguistiques dont, notamment, le traitement du niveau phonologique (e.g., alignement texte/parole) et morphologique (e.g., étiquetage et analyse morphologique des mots) ; l'analyse lexicale (e.g., analyse lexicale distributionnelle) ; l'analyse syntaxique et sémantique de la phrase et du discours ; et l'analyse de la structure du dialogue (e.g., construction de représentations discursives, étiquetage des tours discursifs ; gestion du dialogue). L'analyse lexicale distri-

butionnelle est devenue à l'heure actuelle un champ de recherche très actif au niveau mondial, bien que la composition de ses objets pour obtenir une représentation du contenu d'un constituant de phrase, voire d'une phrase reste un défi théorique majeur (cf. § I.D).

Dans une perspective plus cognitive, le domaine aborde également des thématiques telles que la modélisation du processus d'acquisition de la langue à partir de corpus documentant le développement linguistique des enfants (e.g. le corpus CHILDES : Child Language Data Exchange System) ou encore le développement de modèles ancrant la langue dans l'environnement physique (*grounded language learning*) e.g., de modèles, qui apprennent, à partir de vidéos sous-titrées, la correspondance entre mots et objets.

Enfin d'un point de vue applicatif, le TAL vise le développement de systèmes ou de modules qui permettent un traitement par ordinateur des données langagières et tiennent compte des spécificités du langage humain. Il recouvre à ce titre un champ très large d'applications dont, notamment, la recherche et l'extraction d'information, la veille technologique, la fouille de textes, la correction orthographique et grammaticale, les systèmes de question/réponse, la traduction automatique, les moteurs de dialogue Homme-Machine, la fouille d'opinions, le résumé automatique et la simplification et la génération de textes.

Afin de construire des modèles rendant compte du contexte (visuel, situationnel, épistémique), des différents niveaux de représentation linguistique (phonétique, phonologie, morphologie, syntaxe, sémantique etc.) et de leur inter-relations, le TAL et la LI reposent sur l'utilisation de méthodes symboliques (grammaires computationnelles, approches logiques, systèmes à base de règles), statistiques (apprentissage supervisé, semi-supervisé et non supervisé) et hybrides statistiques/symboliques. Si le développement des corpus d'apprentissage et des méthodes symboliques requiert une expertise linguistique forte (e.g. connaissance de la syntaxe pour la construction d'un corpus

arboré destiné à l'apprentissage d'un analyseur syntaxique), l'utilisation, l'adaptation et l'optimisation des méthodes statistiques aux problématiques langagières exigent par ailleurs, des connaissances poussées dans les domaines de l'algorithmique, des statistiques et de l'apprentissage automatique. Les recherches dans ce domaine sont donc par essence inter-disciplinaires, se situant à l'interface entre informatique, linguistique et modélisation statistique. En outre, l'interface entre TAL/LI et d'autres domaines tels que la psychologie, le traitement du son et de l'image, la robotique et le traitement des connaissances prend actuellement une ampleur renforcée d'une part, parce que les méthodes utilisées (réseaux de neurones, machines à noyaux, champs de Markov conditionnels, etc.) sont de plus en plus communes et d'autre part, parce que l'émergence de nouvelles données conduit naturellement à une synergie entre ces domaines. Ainsi, l'intensification du flux de données sur le web soulève naturellement la question d'un traitement sémantique du texte où les données du web (*linked data*, ontologies, bases de données) seraient les données sémantiques référencées par le texte, de même que la prolifération des documents multimédias soulève celle d'un traitement intégré des données textuelles, audio et vidéo.

Au plan économique et sociétal, le domaine du TAL est porteur de développements industriels importants, pour le développement d'agents conversationnels multimodaux (téléguidage de robots), de l'accès aux connaissances dans les données textuelles et de la préservation du patrimoine linguistique et culturel. En résumé, la demande dans le domaine du TAL ne cesse de croître dans le monde entier. Les offres de postes, dans des centres de recherche publics et privés (Google, Yahoo) comme dans l'industrie, augmentent et les technologies de l'ingénierie linguistique deviennent un enjeu majeur pour traiter des données langagières en isolation et en interaction avec d'autres modalités. Bien que la communauté française du TAL soit structurée autour d'une association forte (ATALA) disposant d'une revue (TAL), d'une conférence nationale (TALN) et d'une liste de diffusion

(LN), elle manque encore d'ouverture internationale comme en témoigne sa faible présence dans les grandes conférences et les grandes revues du domaine.

C. Lexicographie et lexicologie (~ 3%)

Le champ de la lexicologie/lexicographie, conçu comme portant sur l'étude du lexique, indépendamment de son caractère construit (morphologique) ou non, est principalement représenté à l'ATILF (UMR7118), acteur majeur, en France, de la mise à disposition numérique de dictionnaires grâce au CNRTL (Centre National de Ressources Textuelles et Lexicales, <http://www.cnrtl.fr>). Des chercheurs se répartissent dans d'autres laboratoires (CLLE, UMR5263; CRLAO, UMR8563; LLF, UMR7110; L2C2, UMR5304).

Selon les chercheurs, l'objectif peut être de décrire le lexique en synchronie (représentation de connaissances, constitution de lexiques spécialisés ou de langue générale) ou encore dans une perspective diachronique ou comparatiste.

Du point de vue des méthodes, un trait partagé par plus de la moitié des chercheurs de ce champ est le recours aux corpus textuels numérisés, lesquels ont profondément renouvelé la documentation disponible et permettent d'observer des phénomènes auparavant plus difficilement repérables comme les variations diatopiques ou diaphasiques. Deux autres tendances sont la formalisation de la description, qui prend maintenant de plus en plus la forme de bases de données plutôt que de textes, ainsi que la volonté de rendre compte des relations intralexicales au moyen de modèles mathématiques.

Les langues étudiées sont le français et plus généralement les langues romanes, mais aussi le chinois.

V. Histoire des théories linguistiques (~ 3%)

La thématique *Histoire des théories linguistiques* regroupe une dizaine de chercheurs, dont la plupart (8/10) sont regroupés dans le laboratoire HTL (UMR7597). Malgré son effectif réduit, cette thématique couvre trois approches : l'édition critique et/ou la traduction d'œuvres classiques dans l'histoire de la linguistique, comme celles de Boas ou de Benveniste ; la génétique textuelle de textes linguistiques à partir de manuscrits, principalement centrée sur des auteurs francophones, comme Benveniste ; l'histoire des théories linguistiques, couvrant un vaste domaine, allant de la tradition grammaticale sanscrite, aux sémantiques contemporaines.

Le but de l'histoire des théories linguistiques est de mettre en perspective les théories des langues et du langage, relativement à l'histoire de la problématique, à la fois dans le temps profond (tradition grammaticale indienne remontant au IV^e siècle avant J.-C., philosophie du langage et logique médiévales), dans le temps moyen (début de la linguistique au début du XX^e siècle – Saussure –, premiers développements dans le domaine francophone, anglophone, ou russophone – Benveniste, Boas, formalisme russe) et dans le temps récent (linguistique cognitive, approches formelles, grammaires catégorielles).

VI. Démographie de la section

Il a semblé utile de donner quelques indications sur la démographie de la section.

A. Chercheurs

Sur la base des mots-clés fournis par les chercheurs pour qualifier leurs propres thématiques de recherches, une nomenclature en 16 sous-disciplines a été construite. Chacun des 180 chercheurs en activité dans le champ de la section est identifié par son appartenance à une ou plusieurs sous-disciplines (en d'autres termes, le même chercheur peut apparaître dans plusieurs sous-disciplines différentes).

La **figure 1** indique l'implication des chercheurs dans les différentes sous-disciplines ainsi que leur âge, et la **figure 2** leur distribution selon les grades entre ces sous-disciplines.

On notera le pourcentage relativement élevé de chercheurs âgés de plus de 50 ans. À partir de données fournies par le CNRS, il apparaît qu'entre 2003 et 2012, la section 34 a perdu des chercheurs (de 218 à 197), mais pas davantage que les autres sections de l'INSHS. On notera cependant que les départs à la retraite seront nombreux en 2014 et 2015 et que les recrutements annuels (5 postes en 2014) n'arrivent pas à les combler. Au rythme actuel (4 postes annuels de CR), les effectifs de la section passeront entre 2012 et 2016 à 177 et seront de 157 en 2024. Et ceci ne concerne que les chercheurs et pas les ITA.

B. ITA

Les laboratoires en rattachement principal à la section 34 sont 1 UMS, 2 FR, 3 USR et 21 UMR au 31 décembre 2013. Ils comptent 138 ITA CNRS, ce qui représente 83% des effectifs des ITA statutaires déclarés dans Labintel en juillet 2014, le reste étant des personnels universitaires ou autres (28 personnes).

Dans les UMR, le rapport entre les effectifs des permanents/IT, quelque soit leur organisme d'appartenance, et les effectifs des chercheurs et enseignants chercheurs est très variable (cf. Figure 3).

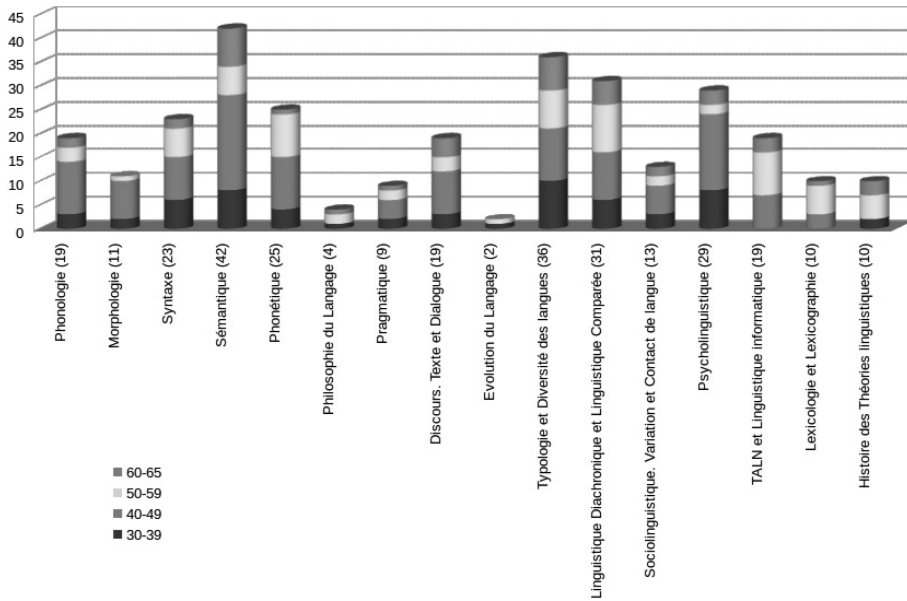


Figure 1

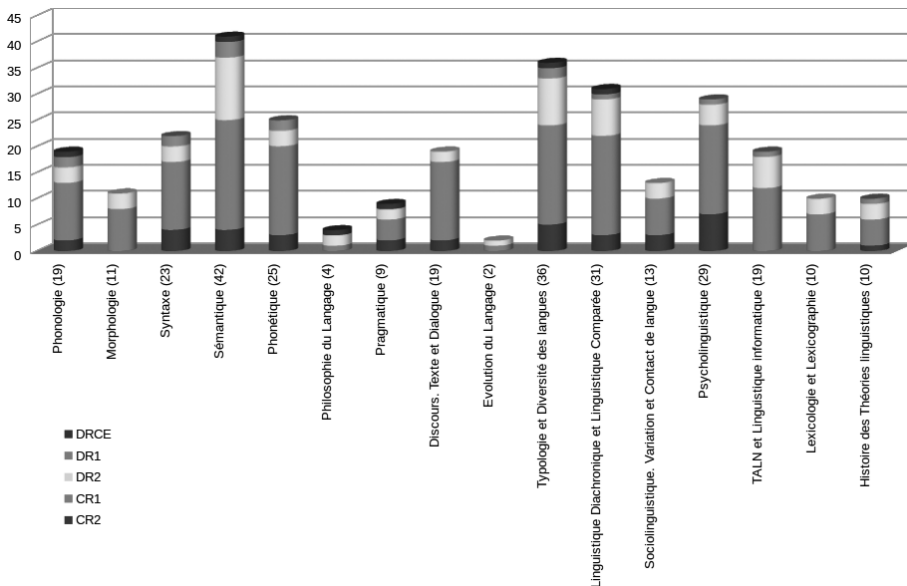


Figure 2

L'âge moyen des 138 ITA CNRS de la section est de 46,8 ans (48,8 ans dans les corps de IR et 48,7% dans celui des IE) ; 56,5% des ITA sont

des femmes. Elles sont minoritaires dans les corps des IR (41,9%), majoritaires dans les corps des AI, des T et des ATR (66,4% en

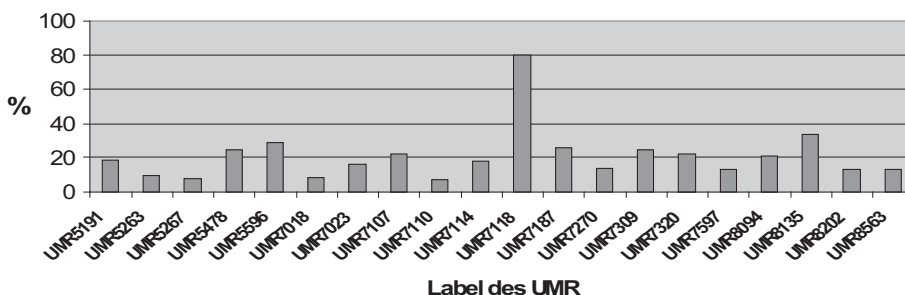


Figure 3.

% Personnel Technique/Chercheurs-Enseignants Chercheurs dans les UMR en rattachement principal.

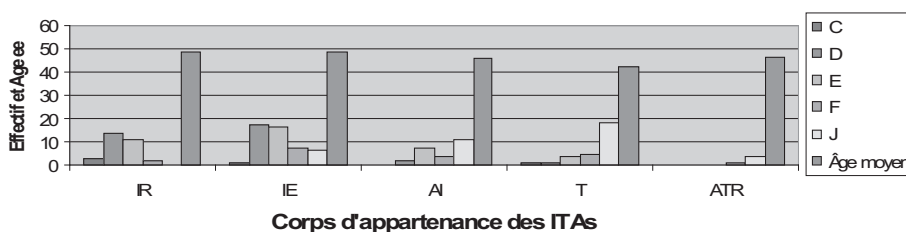


Figure 4.

Effectif et Âge moyen par BAP et par Statut des ITAs des unités en rattachement principal.

moyenne) et sont presque à parité avec les hommes dans le corps des IE (53,2%). Dans les fonctions de soutien à la recherche (BAP C, D, E) les ITA sont en majorité des IR et des IE, alors que dans les fonctions de support (BAP J), ils sont en majorité AI, T ou ATR, la BAP F étant la seule BAP où tous les corps sont représentés (cf. Figure 4).

[BAP: C: Sciences de l'ingénieur; D: Sciences humaines et sociales; E: Informatique, statistique et calcul scientifique; F: Information (documentation); J: Gestion et pilotage]

On notera que la moyenne d'âge des corps IR et IE s'approche de 50 ans, alors que celle des T est plus près de 40 ans. La faiblesse des recrutements actuels, notamment dans les fonctions de soutien à la recherche, est particulièrement inquiétante.

Conclusion

La section 34 est une section dynamique, axée sur l'excellence scientifique, avec une légère majorité de femmes et une répartition relativement égale des sexes dans les différents grades. Néanmoins, la faiblesse des recrutements (en général 4 postes par an) conduit à un vieillissement de la section avec un nombre important de chercheurs âgés de plus de 50 ans. La perspective d'une baisse des recrutements dans les années à venir ne peut qu'aggraver cette tendance à un moment où l'évolution de la discipline et l'orientation vers l'internationalisation des productions devrait au contraire favoriser l'arrivée des jeunes générations. Par ailleurs, la section remarque que les recrutements récents, notamment pour les approches formelles, ont favo-

risé des candidats ayant réalisé leurs doctorats à l'étranger, une évolution largement due à la relative faiblesse des formations en sciences du langage en France. La discipline dans son ensemble profiterait d'une restructuration, notamment au niveau de l'université, favorisant l'ouverture vers les pratiques internationales, en particulier en ce qui concerne les approches formelles.

Enfin, la faiblesse du recrutement ITA, notamment dans les fonctions de soutien à

la recherche, est une autre source d'inquiétude.

Un enjeu majeur pour les évolutions futures est d'équilibrer et de décroïsonner le plus possible les recherches descriptives et cumulatives et les approches expérimentales et formelles. Le rôle de structures transversales comme les fédérations de recherche devrait être principalement orienté sur ce point, crucial pour l'avenir.

Annexe

A. Laboratoires

ALPAGE : Analyse Linguistique Profonde à Grande Échelle (UMRI-001)

ATILF : Analyse et Traitement Informatique de la Langue Française (UMR7118)

BCL : Bases, Corpus, Langage (UMR7320)

CLLE : Cognition, Langues, Langage, Ergonomie (UMR5263)

CRLAO : Centre de Recherches Linguistiques sur l'Asie Orientale (UMR8563)

DDL : Dynamique du Langage (UMR5596)

GIPSA : Grenoble Image, Parole, Signal, Automatique (UMR5216)

GREYC : Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen (UMR6072)

HTL : Histoire des Théories Linguistiques (UMR7597)

ICAR : Interactions, Corpus, Apprentissage, Représentations (UMR5191)

IHPST : Institut d'Histoire et de Philosophie des Sciences et des Techniques (UMR8590)

IJN : institut Jean Nicod (UMR8129)

IKER : Centre de recherches sur la langue et les textes basques (UMR5378)

IREMAM : Institut de Recherches et d'Études sur le Monde Arabe et Musulman (UMR7310)

IRIT : Institut de Recherche en Informatique de Toulouse (UMR5505)

L2C2 : Laboratoire sur le Langage, le Cerveau et la Cognition (UMR5304)

LABRI : Laboratoire Bordelais de Recherches en Informatique (UMR5800)

LACITO : Laboratoire de Langues et Civilisations à Tradition Orale (UMR7107)

LATTICE : Langues, Textes, Traitements Informatiques, Cognition (UMR8014)

LIA : Laboratoire d'Informatique d'Avignon (EA4128)

LIF : Laboratoire d'Informatique Fondamentale de Marseille (UMR7279)

LIMSI : Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (UPR3251)

LINA : Laboratoire d'Informatique de Nantes Atlantique (UMR6241)

LLACAN : Langage, Langues et Cultures d'Afrique Noire (UMR8135)

LLF : Laboratoire de Linguistique Formelle (UMR7110)

LLING : Laboratoire de Linguistique de Nantes (EA3827)

LORIA : Laboratoire lorrain de Recherche en Informatique et ses Applications (UMR7503)

LPL : Laboratoire Parole et Langage (UMR7309)

LPL : Laboratoire de Phonétique et de Phonologie (Paris 3) (UMR7018)

LPP : Laboratoire Psychologie de la Perception (Paris 1) (UMR8242)

LSCP : Laboratoire de Sciences Cognitives et Psycholinguistique (UMR8554)

MoDyCo : Modèles, Dynamiques, Corpus (UMR7114)

SEDYL : Structure et Dynamique des Langues (UMR8202)

SFL : Structures Formelles du Langage (UMR7023)

STL : Savoirs, Textes, Langage (UMR8163)

B. Théories et modèles

CVCV : Consonne Voyelle Consonne Voyelle

DLTAG : Definition List Tag

HPSG : Head-driven Phrase Structure Grammar

LFG : Lexical Functional Grammar

RST : Rhetorical Structure Theory

SDRT : Segmented Discourse Representation Theory

TAL(N) : Traitement Automatique des Langues (Naturelles)

Comité national de la recherche scientifique. « Section 34- Sciences du langage ». *Rapport de conjoncture 2014*, [édition PDF en ligne]. ISBN : 978-2-271-08746-1. Disponible sur : <http://rapports-du-comite-national.cnrs.fr/>