

34

LANGUES, LANGAGE, DISCOURS

Présidente de la section

Jeannine RICHARD-ZAPPELLA

Membres de la section

Pascal AMSILI

Éric BEAUMATIN

Philippe BLACHE

Christian CAVE

Zlatka GUENTCHEVA-DESCLES

Nabil HATHOUT

Alain KIHM

Jean-Marie KLINKENGERG

Sarah LEROY

Philip MILLER

Jean-Luc MINEL

Danièle MONSEUR

Clive PERDUE

Jean-Marie PIERREL

Christian PLANTIN

Laurent SAGART

Paul SIBLOT

Jacqueline VAISSIÈRE

Alain VIAUT

Gérard-Richard WALTER

Ce rapport est le fruit d'une écriture collective associant les membres de la section et des chercheurs et enseignants-chercheurs représentant les communautés scientifiques concernées.

La coordination finale a été assurée par la présidente, Jeannine Richard-Zappella et le secrétaire scientifique de la section, Nabil Hathout.

Les observations présentées ci-dessous ne prétendent pas offrir une description exhaustive des travaux menés dans les laboratoires de la section 34 mais elles visent à présenter un tableau objectif d'une situation marquée de différences sectorielles, où des domaines se développent en interaction avec d'autres disciplines, d'autres cèdent peu ou prou le pas.

La section 34, née du redécoupage des sections du comité national en 1991, regroupe des disciplines différentes consacrées aux sciences du langage, aux systèmes sémiotiques, etc. Son nouvel intitulé «Langues, Langage, Discours» confère une visibilité plus forte au périmètre et au cœur de la discipline. Science autonome et diverse, la linguistique étudie le langage humain sous tous ses aspects. Parler du langage humain renvoie au constat suivant : les hommes parlent et sont les seuls à le faire. L'étude du langage est donc un des moyens privilégiés pour étudier ce qui est une capacité de l'espèce humaine. Quant aux langues, il s'agit essentiellement de déterminer

les propriétés qui les caractérisent par rapport à d'autres systèmes existants ou envisageables. En d'autres termes, on cherche à discriminer quelles propriétés sont universelles, quelles sont particulières.

Enfin, l'objectif des sciences du langage est de décrire comment l'agencement de certaines marques matérielles permet aux hommes de produire du sens, production socialement partagée.

1 – QUELQUES REMARQUES LIMINAIRES

Les domaines de la section sont multiples, certains traditionnels, d'autres plus récents. Ses objets premiers sont :

- la diversité des langues naturelles ;
- la fonction langage comme capacité spécifique à l'espèce humaine ;
- les interactions langagières.

Ses principaux objets d'étude et champs traditionnels sont les suivants :

- phonétique ;
- phonologie ;
- morphologie ;
- syntaxe ;
- lexicologie ;
- lexicographie ;
- sémantique ;
- pragmatique ;
- histoire ;
- description et typologie des langues ;
- étude des textes et des discours ;
- psycholinguistique et mécanismes cognitifs ;
- sociolinguistique ;

- ethnolinguistique ;
- traitement automatique des langues et ingénierie linguistique.

De par sa composition, la section 34 est ouverte à toutes les sous-disciplines des SDL et le carrefour de points de vue théoriques multiples. Cette plurivocité qui s'explique tant par la pluralité des domaines que par la multiplicité des langues étudiées ne peut être que féconde même si elle est aussi porteuse de tensions. En résumé, tant du fait de la pluralité des domaines que de la multiplicité des langues étudiées, les sciences du langage, en France comme ailleurs, se caractérisent par une grande diversité d'approches complémentaires, aussi bien théoriques qu'empiriques. L'absence d'unification théorique est caractéristique de l'état actuel du développement des sciences du langage, et la diversité des approches doit absolument être préservée.

On soulignera également quelques caractéristiques spécifiques au champ qu'un rapport de conjoncture se doit de mentionner. Il existe des différences notoires de structuration des disciplines entre le CNRS et les CNU. C'est ainsi que le regroupement de chercheurs et des unités de recherche qui correspond à la section 34 n'a pas d'équivalent dans les structures universitaires ; la section 34 couvre beaucoup plus de disciplines que la section 7 du CNU « Sciences du langage : linguistique et phonétique générale ». À cela s'ajoute une structuration de l'enseignement supérieur par grandes aires linguistiques, structuration qui ne permet pas de donner au champ toute la visibilité qu'il mérite. Ainsi de nombreux enseignants-chercheurs relèvent pour leur recrutement et leur évaluation des sections du CNU autres que la section 7 mais dépendent de la section 34 du Comité national pour leur activité de recherche.

Si le CNRS a joué et continue de jouer un rôle fondamental dans la recherche linguistique sur des langues peu ou pas représentées à l'université, un déséquilibre perdure entre le poids du CNRS dans la recherche et la formation sur les langues de « terrain » et le rôle plus que marginal de l'enseignement supérieur dans ce secteur. Ce dysfonctionnement n'est

pas sans conséquence : vieillissement et départ à la retraite des chercheurs recrutés dans les années 70, souvent seuls spécialistes de leur domaine, qui ne se voient pas remplacés par les flux de recrutements actuels.

En outre, force est de constater que ces disparités, que l'absence d'un cursus commun, d'un socle commun de connaissances clairement identifiable au niveau de la formation universitaire ont permis au Ministère de tutelle de s'engouffrer dans la brèche et de remettre en cause les SDL, au cours de l'année 2004, lors de la campagne d'habilitation des diplômés dans la réforme LMD. Les arguments avancés reposaient sur une certaine méconnaissance de la discipline, de son intérêt et de son apport scientifique dans la modernisation des enseignements dans les cursus et les concours qui conduisent aux métiers de l'enseignement. Il est de ce fait important de renforcer l'appartenance disciplinaire, de la rendre suffisamment visible pour construire une identité disciplinaire forte.

De plus, le cadre disciplinaire français est atypique au niveau européen. Les formulations et les aires d'affiliation varient selon les pays ; les sciences du langage peuvent être ainsi rattachées aux arts et lettres, aux sciences humaines et sociales, aux sciences de l'information et de la communication.

Pour autant toute nouvelle recomposition qui viendrait déjouer les anticipations des chercheurs, des universitaires, devrait être solidement justifiée pour éviter que la perte soit supérieure au gain.

2 – PANORAMA DE LA DISCIPLINE

2.1 UNE INTERDISCIPLINARITÉ DOUBLE

La place centrale qu'occupe le langage dans les activités humaines confère à la section

un caractère interdisciplinaire fort. Pour l'ensemble de la communauté cette interdisciplinarité est à la fois une évidence et une règle mais elle est également et surtout la garantie de la visibilité et vitalité de la discipline. Les interactions portant sur des sujets à forte demande sociale où des découvertes importantes sont possibles à court terme doivent être encouragées et soutenues. Elles concernent en particulier :

- la psycholinguistique et les sciences cognitives ;
- le traitement automatique des langues, l'ingénierie linguistique et les sciences et technologies de l'information ;
- la linguistique historique, l'archéologie et la génétique des populations ;
- la philosophie, la logique ;
- l'ethnologie ;
- les sciences des textes, etc.

Néanmoins si la linguistique fut dans les années 70 la discipline phare des SHS, elle a souffert ces dernières années de la fascination qu'opèrent les sciences dites dures, les courants interdisciplinaires qui entretiennent des liens plus étroits avec des disciplines des sciences plus dures et avec d'autres départements. D'une manière générale les chantiers interdisciplinaires internes occupent un vaste espace qui s'explique, faut-il le rappeler, par l'importance du langage comme activité sociale et comme pratique culturelle. Si ce message semble à nouveau bien passer au niveau des instances décisionnelles, il est souhaitable que cette interdisciplinarité se voit soutenue notamment en augmentant le nombre de postes au concours, à condition cependant que cette politique ne se fasse pas au détriment du cœur de la discipline.

Quant à l'interdisciplinarité générale, elle fait l'objet d'une très grande attention de la part de la direction ; depuis 15 ans la section 34 a construit des liens étroits avec les disciplines relevant de différents départements, notamment avec les chercheurs des sections 29 et 07. Ces interactions concernent d'une part la

psycholinguistique et les sciences cognitives, d'autre part le traitement automatique des langues et l'ingénierie linguistique.

Malgré les difficultés, voire les clivages liés notamment à des découpages disciplinaires différents entre le CNRS et le monde universitaire – la psycholinguistique et la neurolinguistique relèvent de la section 29 et des Sciences de la Vie alors qu'elles sont intégrées aux sciences humaines et sociales à l'Université – la politique de recrutements croisés mise en place ces dernières années a permis une collaboration féconde entre les deux communautés. Parallèlement, on assiste à la naissance de coopérations entre des chercheurs relevant de la section 34 et les neurosciences intégratives, notamment avec la présence de l'imagerie cérébrale hors de son champ d'origine.

De manière plus générale, la section 34 est très impliquée dans le champ des sciences cognitives. La politique très volontariste des tutelles permet un développement fort du domaine et l'implication de différents courants de la recherche, essentiellement de deux en ce qui concerne la section 34. Si le premier rassemble des chercheurs qui se réclament de la philosophie de l'esprit anglo-saxonne, des grammaires formelles d'inspiration chomskienne et des différentes hypothèses innéistes sur les propriétés universelles des langues, le second s'inscrit davantage dans la tradition de la recherche linguistique européenne et adopte le point de vue des grammaires cognitives. On ne peut que se féliciter de cette diversité des approches qui se révèle particulièrement féconde.

On notera également que les sciences et technologies de l'information constituent l'autre grand domaine d'interface de la section. Si ces coopérations ont connu dans les années 50 un essor, elles bénéficient aujourd'hui des apports du traitement de la parole, de la recherche d'information, de l'acquisition de connaissances, des interactions homme-machine, de la traduction automatique relayée ensuite par le traitement automatique des langues et de la linguistique informatique.

Néanmoins si cette interdisciplinarité est un atout fort pour la section 34, elle ne doit pas

cependant prendre le pas sur le cœur de la discipline. Une politique volontariste de fléchage de postes interdépartements a abouti à une part importante des recrutements de candidats se situant aux interfaces. Si un tel développement constitue une source de richesses il ne doit pas s'opérer au détriment d'autres domaines des sciences du langage. On citera trois exemples qui, à des titres différents, méritent l'attention : la « linguistique des langues », la « linguistique du langage », et le vaste domaine de la textualité, du discours et des pratiques langagières.

2.2 TENDANCES PAR GRANDS DOMAINES

Linguistique historique, archéologie et génétique des populations

Formation des familles de langues et peuplement du monde

Alors que l'idée d'un peuplement du monde par notre espèce au cours des derniers 150 000 ans gagne en crédibilité et que les généticiens des populations construisent des hypothèses détaillées sur les routes et les étapes de ce peuplement, les linguistes historiques apportent une contribution décisive à ces modèles pour la période « récente », celle de la formation des grandes familles de langues modernes au cours des 10 000 dernières années. Ces collaborations, menées aujourd'hui principalement en Europe, et portant sur l'Afrique, l'Eurasie et le Pacifique, font aussi intervenir des archéologues, notamment dans le cadre des théories de Bellwood et Renfrew sur le rôle des transitions à l'agriculture dans la formation de diverses familles de langue. Il faut par ailleurs noter que le contact avec la génétique et la biologie évolutive a permis d'introduire en linguistique historique des concepts et des méthodes de classification venus de la cladistique, qui ont fortement stimulé les travaux des linguistes sur la généalogie des langues et abouti à de nouvelles propositions qui sont en train d'être testées.

Origine du langage

Signalons la réémergence récente du problème de l'origine du langage (programme OHLL notamment), qui constitue actuellement un domaine très interdisciplinaire en pleine expansion, mettant en jeu la linguistique, l'anthropologie physique, l'ethnographie, la paléodémographie, la paléoclimatique, l'archéologie, la psychologie cognitive et la primatologie.

Linguistique des langues

Typologie

Autrefois orientée principalement vers la description et l'histoire des langues, la linguistique des langues a connu depuis une quinzaine d'années l'apparition de d'une linguistique plus centrée d'une part sur les problèmes de typologie et la recherche d'universaux, et d'autre part sur les problèmes cognitifs posés par la diversité des langues. Cette communauté est de plus en plus ouverte vers l'extérieur, et elle occupe une place de choix dans la recherche internationale. En témoignent par exemple les liens internationaux forts que les linguistes de la fédération TUL entretiennent avec les typologues hors de France. Ces efforts doivent être encouragés et soutenus.

Langues en danger, archivage

Autre élément d'évolution, la prise de conscience de plus en plus aiguë de la disparition imminente d'un grand nombre des langues parlées aujourd'hui, dont peu sont bien décrites ; c'est le devoir des linguistes de notre génération de décrire ces langues afin d'en préserver ce qui peut l'être (phonétique/phonologie, lexique, grammaire, textes) pour les générations futures, tant pour ce qu'elles peuvent nous apprendre sur la capacité humaine de langage que pour les informations qu'elles recèlent sur l'histoire non écrite des peuples

qui les parlent. Il est évident que le CNRS a une mission particulière en ce qui concerne les langues parlées sur le territoire national, en particulier dans le Pacifique et en Guyane. Un effort international est en cours (UNESCO, Hans Rausing Fund/SOAS, Volkswagen/MPI) pour financer le recueil des données sur le terrain, la production de grammaires et de dictionnaires, ainsi que pour préserver sous forme audio et vidéo des textes oraux accompagnés de transcriptions et de traductions. Les linguistes français prennent une part active à ce mouvement : ils souhaitent que le CNRS accompagne leurs efforts en finançant leurs missions de terrain, et en accompagnant leurs efforts d'archivage des données recueillies.

La Direction scientifique pourrait envisager d'inciter des fondations privées ou de grandes entreprises à prendre en charge le coût financier de ces efforts, tout en conservant le contrôle scientifique de l'opération, comme en Allemagne (fondation Volkswagen sous la direction du MPI) ou en Angleterre (Fondation caritative Hans Rausing, administrée par la School of Oriental and African Languages de Londres).

On signalera toutefois un élément préoccupant, lié au vieillissement des chercheurs recrutés dans les années 1970 et 1980. Les vagues de départs en retraite en cours privent ce champ de chercheurs de qualité, souvent seuls spécialistes de leur domaine, sans que les flux de recrutement actuels permettent leur remplacement dans des conditions satisfaisantes. La linguistique des langues risque à court terme de se transformer en secteur sinistré du CNRS.

Linguistique historique et comparative

En partie de tradition française, elle a dans la deuxième moitié du xx^e siècle, et singulièrement dans les 20 dernières années, subi une profonde évolution du fait du développement de nouvelles théories sur le contact des langues et la créolisation (Thomason-Kaufman, Bickerton), le changement linguistique (Labov, Milroy) et la grammaticalisation (Heine), et se

trouve de ce fait dans un certain état d'effervescence, avec des débats animés sur la validité du modèle généalogique de parenté des langues et les méthodes de leur classification (avec des apports théoriques venus tant de la biologie que de la théorie de la complexité).

La linguistique du langage

La prosodie et ses interfaces

Le premier domaine est celui de la prosodie, c'est-à-dire l'ensemble des phénomènes liés à l'utilisation signifiante des variations de hauteur, d'intensité et de durée du signal sonore signifiant. Cette définition vise à faire voir que la prosodie, dont l'objet empirique est la courbe intonative, fonctionne à deux niveaux de signification simultanément : d'une part, la signification associée au signal sonore « plat », par exemple ce que veulent dire (dénotent) les éléments *Chloë, a nourri, le* et *chat* et leurs combinaisons dans l'énoncé-type *Chloë a nourri le chat*; d'autre part, la signification de la courbe intonative particulière tracée par les valeurs des paramètres prosodiques aux points successifs de la séquence segmentale réalisée [kloeanurilʃa].

La prosodie se saisit donc par définition au niveau d'interfaces, qui sont au moins au nombre de trois :

- l'interface prosodie-syntaxe : l'analyse des (non-)correspondances entre les points de transition de la courbe intonative et les frontières des groupes syntaxiques ;

- l'interface prosodie-sémantique : l'analyse du marquage prosodique des éléments de l'énoncé situés dans la portée d'un opérateur lié au focus (adverbes du type de *seulement*, particules focalisatrices, constructions clivées, etc.) ;

- l'interface prosodie-pragmatique : l'analyse des relations entre l'inventaire des courbes intonatives et les inflexions de la structure informationnelle (engagement du locuteur dans ce qu'il énonce, posture vis-à-vis de l'interlocuteur, etc.).

L'enjeu théorique de l'étude est donc fort, puisque la prosodie se connecte aux trois composantes interprétatives du dispositif linguistique, si bien qu'elle occupe une place centrale dans l'usage interactif du langage oral. Elle a des analogues (et/ou des ancêtres) probables dans les langages animaux, hors de l'espèce humaine. On ne lui voit guère d'équivalent, en revanche, dans les langages logiques et informatiques. Son fondement empirique est également solide, formé qu'il est d'analyses instrumentales très précises. Enfin, on est surpris de l'ampleur de ce qui reste à découvrir, en matière de faits et de formalisations, d'un phénomène aussi essentiel et aussi quotidien.

Phonologie de laboratoire

Les recherches rassemblées sous le terme de phonologie de laboratoire visent à aborder dans une perspective expérimentale les principales questions qui se posent à la phonologie d'aujourd'hui : émergence et évolution des systèmes phonologiques, relations avec les autres composantes de la grammaire, universaux phonologiques et variabilité inter/intra-langue, processus d'acquisition chez l'enfant, forme et fonction des représentations phonologiques dans la production et la compréhension du langage oral, etc. L'accent est placé sur les relations qui s'établissent entre phonologie, sciences de la parole, et sciences cognitives. Les méthodes employées en phonologie de laboratoire sont pour une partie d'entre elles empruntées à la psychologie expérimentale, aux sciences de la vie et aux sciences physiques.

Une importance majeure est accordée aux procédures expérimentales de validation des hypothèses. Par exemple, on entreprend de reproduire en laboratoire les mécanismes que l'on suppose à l'œuvre dans l'évolution des systèmes phonologiques (sound change). La phonologie de laboratoire fait aussi largement recours aux modèles computationnels (modèles connexionnistes, systèmes artificiels multi-agents, etc.) dans le but de simuler la manière dont les unités phonologiques peu-

vent émerger dans la mise en relation entre son et sens. On cherche par ailleurs à caractériser avec précision les contraintes articulatoires, acoustiques, perceptives et cognitives susceptibles de s'exercer sur l'architecture des systèmes phonologiques, à travers des études instrumentales à cheval entre la phonologie, la phonétique et la psychologie cognitive. Le rôle susceptible d'être joué par l'usage sur le niveau phonologique donne lieu à des investigations approfondies avec l'avènement des théories phonologiques probabilistes par exemple. Les mécanismes mis en œuvre dans l'acquisition de L1/L2 ainsi que les désordres langagiers sont également au centre de la phonologie de laboratoire. La phonologie de laboratoire se place ainsi au carrefour entre plusieurs disciplines afin de mieux caractériser la façon dont le locuteur et l'auditeur se représentent la forme sonore du langage.

La phonologie de laboratoire constitue aujourd'hui un champ en plein essor, comme en témoignent les nombreuses publications dont ces travaux font aujourd'hui l'objet, ou le succès de la conférence LabPhon, dont la dernière édition s'est tenue à Paris en 2006.

La morphologie

Plus encore peut-être que la prosodie, les phénomènes morphologiques paraissent entièrement spécifiques du langage humain. D'un point de vue strictement fonctionnel et d'ingénierie (*design*), un langage n'a, semble-t-il, besoin de rien de plus que d'un lexique (avec ou sans double articulation selon la quantité d'items requise), d'une syntaxe et d'un ensemble d'opérateurs logiques (conjonction, disjonction, quantificateurs, etc.). Or, aucune langue naturelle ne s'aligne sur ce schéma idéal, si bien qu'on est en droit d'inférer que le langage humain ne s'y conforme pas.

D'une part, les items lexicaux ou lexèmes naturels sont affectés à des catégories (noms, verbes, etc.) qui ne correspondent que partiellement aux catégories ontologiques élaborées (entités, événements, etc.). Cette catégorisation est souvent corrélée à des différences systéma-

tiques dans la forme phonologique des lexèmes.

Il est en outre fréquent – peu de langues naturelles sont tout à fait dépourvues de tels phénomènes – que cette forme varie, de façon différentielle selon la catégorie, en tout ou (le plus souvent) en partie, selon la fonction grammaticale (sujet, complément, adjectif) du lexème; et/ou selon qu'un lexème de sens constant est affecté d'une modification de sa dénotation (*voir Il pleuvait vs. Il pleut, cheval vs. chevaux*); et/ou selon qu'il entretient une certaine relation avec un autre lexème, lui-même soumis à variation (accord). On délimite ainsi ce qu'on est convenu d'appeler la morphologie flexionnelle, dont le mode d'organisation est le paradigme.

Il est d'autre part constant que les variations de forme servent à constituer des familles transcategorielles de lexèmes unifiées par un fond sémantique commun (*voir informer, information, informatique, etc.*). On parle ici de la morphologie dite dérivationnelle ou constructionnelle.

La complexité de ces procédés est souvent très grande, tant par le nombre de formes associées qu'ils génèrent que par l'absence fréquente de régularités évidentes permettant de prédire avec sûreté telle forme à partir de telle autre. On ne peut donc éviter de se poser la question de leur nécessité en termes de meilleure adaptation aux fonctions du système – au sens où la morphologie des organismes vivants est due en grande partie à de telles contraintes adaptatives.

Ce faisant, on se gardera de poser en a priori que les phénomènes morphologiques sont dysfonctionnels en raison de leur complexité même. La mesure de celle-ci ne va pas de soi. Il est clair que l'effet des variations de forme pourrait être obtenu (et l'est à l'occasion) par une combinatoire d'unités discrètes syntaxiquement ordonnées, sur le modèle des formules de la logique et des langages de programmation. Mais cela serait-il plus simple, en quel sens et pour qui? Les langues naturelles ne semblent-elles pas conforter l'apparent paradoxe – si bien mis en lumière par les tra-

vaux de Claude Lévi-Strauss, pour ne citer que lui – que, pour les humains, c'est le complexe qui est le plus simple?

Par ces réflexions trop sommaires, nous ne voulons que pointer l'importance des études morphologiques. Loin d'être une ex-croissance, voire un embarras qu'il vaudrait mieux supprimer (comme le font les langues artificielles qui se veulent « parfaites », mais pas, c'est symptomatique, les « naturalistes » comme l'espéranto), la morphologie constitue une pièce centrale et distinctive du dispositif linguistique humain alias faculté de langage : centrale par les interfaces qui la connectent aux autres pièces du dispositif (phonologie, lexicale, syntaxe, sémantique); distinctive parce que à nouveau, comme la prosodie, sans équivalent évident dans les autres systèmes de connaissance et d'expression.

L'analyse formalisée des procédés morphologiques des langues naturelles, comparés entre eux et dans une perspective inter-systémique (comparaison avec les langages construits, les « mythologies », les écritures), rapportée aux études sur l'acquisition, apparaît ainsi comme une source privilégiée de découvertes aptes à enrichir notre connaissance du fonctionnement de l'esprit humain – ce en quoi la linguistique est en effet une science humaine et ce sans quoi elle n'aurait guère d'intérêt.

Les grammaires de référence

On a vu paraître ces dernières années une série d'excellentes grammaires de référence de plusieurs langues romanes : catalan, espagnol, italien, portugais, roumain. Le français ne sera bientôt plus en reste puisqu'actuellement une grammaire est en construction, sous l'égide de l'Institut de Linguistique Française, l'une des deux fédérations de référence pour notre section.

Ces grammaires ont pour caractéristique de concilier la technicité et un certain degré de vulgarisation. Rédigées par des équipes de linguistes locuteurs natifs et spécialistes de chacun des domaines inclus (phonologie, mor-

phologie, syntaxe, sémantique), elles requièrent une « culture » linguistique poussée pour être utilisées. Ce ne sont pas des ouvrages « grand public ». Leur lectorat naturel est sans doute formé de linguistes de toutes spécialités, ayant besoin pour leurs recherches d'informations fiables et précises sur les langues décrites. Mais ce ne sont pas non plus des ouvrages adressés aux seuls linguistes : ils sont aussi destinés à quiconque a besoin de ces mêmes informations pour des raisons pratiques, l'enseignement en particulier. Cet accessibilité est assurée par le fait que, quoique techniques, ces grammaires n'utilisent pas des formalismes très élaborés, entièrement opaques pour les profanes.

En peu de mots, ces grammaires ont pour ambition de rassembler et de rendre disponible la somme des connaissances cumulées sur une langue à l'époque de leur rédaction. Ces connaissances sont présentées sous une forme telle que, sans être formalisée au point de n'être lisible que des seuls spécialistes, sa rigueur permet une formalisation immédiate à fin de recherches.

Les projets de ce type nous paraissent très importants. Leur utilité pour la communauté des linguistes est indéniable. Qui parmi nous ne souhaite avoir à sa disposition des sources de données et d'exemples fiables sur des langues qu'il ne connaît pas ou mal? Mais leur intérêt est plus vaste. 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.

Le CNRS doit soutenir la production de tels outils.

Psycholinguistique et sciences cognitives

L'enjeu pour ce domaine de recherche interdisciplinaire est de mieux comprendre les différents facteurs psychologiques et neurobiologiques qui rendent l'être humain capable d'acquérir et d'utiliser le langage.

Plus précisément, on cherche à décrire les processus mentaux qui sous-tendent le traitement (l'acquisition, la production, la perception et la compréhension) du langage et les liens qui s'établissent dans ces processus entre langage et pensée. Par rapport à ce dernier point, de nombreux psycholinguistes cherchent à définir la part de l'universel par rapport aux spécificités des langues individuelles en analysant les interactions entre le développement cognitif et le développement langagier de l'enfant monolingue ou bilingue, ou bien les interactions entre la conceptualisation d'un message à produire, et la formulation de ce message dans une langue donnée, processus que D. Slobin a baptisé « *thinking for speaking* ». Les méthodes d'investigation de ces différents processus reflètent bien l'interdisciplinarité du domaine, empruntant à la linguistique de corpus, aux paradigmes de la psychologie expérimentale ainsi qu'aux méthodes d'imagerie utilisés dans les neurosciences cognitives. Une façon fructueuse de cerner les différents aspects du traitement du langage est l'approche comparative : on compare les performances de locuteurs de différentes langues maternelles face à une tâche identique ; en gardant constante la tâche et la langue, on compare les performances de locuteurs de différents âges, ou bien les comportements de locuteurs tout-venant et de locuteurs souffrant de différents dysfonctionnements : aphasiques, dysphasiques, etc.

La psycholinguistique est une discipline très jeune en mutation rapide et constante. On peut évoquer deux raisons pour expliquer ceci : la rapidité des progrès technologiques (imagerie cérébrale), mais surtout à la prise en compte progressive des variations importantes attestées dans la complexité formelle des structures de différentes langues, ce qui permet de mieux cerner la relation entre langue et cognition. De ce point de vue, l'interaction entre psycholinguistique et les recherches menées sur la typologie linguistique et les universaux du langage s'avère prometteuse.

Usages langagiers

Textualité, discours, pratiques langagières

Au même titre que les enjeux cognitifs, la description des langues et la typologie, le domaine de la textualité, du discours et des pratiques langagières constitue l'un des axes prioritaires du département dans les orientations stratégiques publiées en 2000. Ce texte n'a rien perdu de son actualité : « Les sciences humaines et sociales attendent de la linguistique qu'elle développe des méthodes d'analyse permettant de comprendre ce qui se joue au niveau [des productions langagières]. Les productions monologiques et les interactions verbales doivent faire l'objet d'une approche théorique et méthodologique unifiée [...] Les productions écrites et orales doivent être appréhendées en respectant leurs variabilités et leurs relatives spécificités ». Il est souhaitable que ce programme soit réellement mis en œuvre, pour permettre aux recherches sur les textes, les discours et les pratiques langagières d'atteindre une masse critique, intégrant mieux l'étude de la production verbale, orale et écrite, longtemps négligée par la recherche linguistique. C'est un domaine où la demande sociale d'une expertise linguistique innovante est forte.

Interactions langagières

La parole en interaction s'inscrit dans une vision du langage comme lieu de communication et d'action ; elle constitue la forme première d'existence sociale du langage. Elle joue un rôle constitutif dans l'organisation des relations sociales (ordinaires, professionnelles, institutionnelles, etc.) ; elle est notamment le lieu de l'apprentissage des connaissances et de leur mise en œuvre, comme de l'évolution des langues et des formes de discours. Les machines sont maintenant partie prenante de ces processus. Les recherches sur les interactions communicatives Humain-Humain relèvent d'une linguistique de terrain ont joué un rôle pionnier dans la constitution d'une communauté

d'étude des usages. Elles sont fondées sur la constitution de corpus de données spécifiques (enregistrements audio/vidéo, transcriptions) de parole en interaction recueillies dans des conditions «écologiques». La constitution de tels corpus relève de pratiques professionnelles de recherche qui intègrent des savoir-faire techniques, méthodologiques et analytiques tenant compte de leurs spécificités (approche ethnographique; multimodalité; transcriptions; contraintes juridiques, déontologiques et éthiques sur le recueil, le stockage et la diffusion de ces données). Les analyses empiriques menées sur ces corpus ont abouti à la constitution de corps de savoirs constitués, portant non seulement sur les divers niveaux descriptifs de la langue, mais également sur l'ancrage multimodal de l'usage du langage en situation et dans l'action, au sein de communautés de parole.

Les interactions verbales, médiatisées par machine ou non, constituent un paradigme dans lequel de nombreux domaines (linguistiques et autres), ont émergé et évolué pendant les derniers dix années. Le développement de ces recherches dépend maintenant, de façon cruciale, de la structuration de ces corpus en banques de données partagées et outillées (requêtes, data-mining) (*voir* § Corpus et outils linguistiques). Les interactions langagières Humains-Machines et entre les humains à travers les machines font l'objet de nombreux travaux récents qui visent à prendre en compte la multimodalité et la cognition située et distribuée. Les avancées dans ce domaine amènent à prendre en compte des modèles comme les modèles de la langue, de l'interaction Humain-Humain, de l'acquisition du langage, du fonctionnement cognitif et langagier en sciences cognitives, des protocoles sociaux qui structurent les conversations. Construire des modèles computationnels qui puissent en rendre compte, mais aussi les évaluer et les faire évoluer est un objectif important des prochaines années. Les objectifs socio-économiques attendus sont importants: instrumenter la communication humaine (mémoire, gestion du temps), faciliter les interactions humains-machines (expression, ergonomie, esthétique), définir le statut social des machines dans l'inte-

raction (substituts, mandataires, intermédiaires), les droits et devoirs qui s'y attachent en tant que «parlant».

La collaboration pluridisciplinaire est essentielle pour mener à bien ses projets, tant pour l'analyse de corpus Humain-Humain dans des conditions nouvelles que pour l'évolution des artefacts langagiers. Elle permet de croiser l'analyse des corpus et des résultats et de construire des modèles reconnus tant par les informaticiens que par les chercheurs en sciences humaines.

Traitement automatique des langues, ingénierie linguistique et sciences et technologies de l'information

TAL

Le traitement automatique des langues (TAL) et la linguistique de corpus sont devenus, au cours des dernières années, des domaines-clés pour répondre aux besoins de notre société en terme d'analyse et d'exploitation de gisements d'information, le plus souvent sous forme textuelle. Les besoins en termes descriptifs et formels se sont ainsi développés, couvrant aussi bien les aspects lexicaux, syntaxiques, sémantiques, ou pragmatiques. Il devient ainsi nécessaire de proposer des modèles s'appuyant sur une double validation, explicative d'un point de vue linguistique, opératoire d'un point de vue informatique. Par ailleurs la disponibilité de ressources textuelles électroniques de grandes tailles (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. Aujourd'hui, traitement automatique des langues et linguistique de corpus structurent un nouveau champ disciplinaire aux finalités multiples, en particulier :

- modélisation de la langue, de sa structure et de son usage conduisant la linguistique à des exigences d'opérationnalité effective sur les

formes d'usage de la langue, par opposition aux exemples construits ;

– mise en place d'applications concrètes : indexation et accès à l'information, résumé de textes, extraction de connaissances, dialogue homme-machine, par exemple.

Par ailleurs un des enjeux majeurs du TAL réside maintenant dans le traitement de matériel linguistique non canonique : langue parlée, textes tout-venant, courriels, etc. Les modèles théoriques joueront dans cette perspective un rôle déterminant : ils offrent en effet la possibilité d'une description précise de chacun des domaines abordés par la linguistique (phonétique, morphologie, syntaxe, pragmatique, etc.) et permettent de plus d'en expliquer les interactions de façon cohérente et systématique. Il devient ainsi possible d'aborder des domaines d'investigations nouveaux pour le TAL comme l'analyse du discours ou des interactions. 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 d'une information multimodale. Les grands domaines d'application du TAL relevant du traitement de l'information ou de l'interaction homme-machine trouveront ainsi des bases théoriques et descriptives à partir desquelles développer de nouvelles orientations.

Les ressources

Ces objectifs, autant théoriques qu'appliqués, nécessitent des méthodes et des outils spécifiques. La question des ressources est en particulier cruciale.

Le premier type de ressources concerne les corpus textuels et les corpus d'exemples. Leur rôle est en effet central pour le développement et l'évaluation de modèles opératoires ou de théories linguistiques représentatives de l'usage effectif de la langue. Il s'agit le plus souvent de faire émerger des invariants ou, au contraire, des comportements particuliers

d'entités linguistiques. Le web constitue bien entendu aujourd'hui une source importante d'information, mais de qualité souvent discutable. Il est donc indispensable de disposer également de ressources de qualité validées.

Le second type de ressources concerne les dictionnaires et les lexiques. Aucun traitement automatique de la langue ne peut se passer du niveau lexical, et la disponibilité de ressources de ce type est indispensable pour la plupart des traitements. Là encore les besoins sont très divers dans un contexte mono ou multilingue : dictionnaires spécialisés et dictionnaires généraux de langue, lexiques techniques ou bases terminologiques, par exemple. Il est donc important de développer et de partager des ressources de ce type, comme par exemple la version informatisée du Trésor de la Langue Française (www.tlfi.fr) et sa dérivation sous forme d'un lexique ouvert des formes fléchies du français (540 000 formes issues de 68 000 lemmes : www.atilf.fr/morphalou).

Le troisième type de ressources, complément des deux précédents, concerne les outils de traitement de la langue. Deux types d'outils méritent une attention toute particulière :

– i) Les outils de gestion et d'exploitation des ressources textuelles, lexicales ou dictionnaires ;

– ii) Les outils de base de traitement de la langue permettant en particulier la segmentation, la lemmatisation, la conjugaison, l'étiquetage morpho-syntaxique, le parenthésage, etc.

La linguistique de corpus

La modélisation, qui est au cœur de toute activité de recherche en linguistique, a pour but premier, dans notre domaine, de proposer ou parfaire des théories linguistiques qui soient tout à la fois opératoires et valides d'un point de vue de l'usage de la langue. C'est ainsi qu'aujourd'hui l'accent est mis de façon plus forte sur l'usage effectif de la langue tel qu'il peut être analysé à travers l'exploitation de vastes corpus textuels représentatifs d'un domaine applicatif ou plus généralement d'un

usage nouveau tel qu'il apparaît sur le Web. Ce courant fut initié dès la fin des années 60 en lexicographie notamment à travers le projet de dictionnaire du Trésor de la Langue Française, aujourd'hui disponible sous forme électronique (www.tlfi.fr). Cette base de données textuelles, enrichie et mise à jour, a donné naissance à Frantext (www.atilf.fr/frantext), sans aucun doute le plus grand corpus diachronique sur la langue française (220 millions d'occurrences de mots, soit plus de 1,5 milliards de caractères), support aujourd'hui de nombreuses recherches en linguistique de corpus.

Au cours des dernières années, parmi les résultats les plus remarquables, il convient de noter ceux obtenus aux niveaux lexical et terminologique (création de lexiques multilingues), morphologique et morphosyntaxique (en particulier étiquetage morphosyntaxique de textes), ou syntaxique (analyse syntaxique robuste d'énoncés) ainsi que leur exploitation soit comme prétraitement pour un accès au contenu de données textuelles, soit comme aide à l'enseignement des langues assistée par ordinateur.

La linguistique de corpus et le TAL sont ainsi devenus des domaines-clés répondant aux besoins actuels de notre société de l'information. Les principales activités visées sont les industries de la langue, l'édition numérique, la veille technologique, mais aussi la gestion de patrimoines scientifiques et techniques (IST), industriels (mémoire d'entreprise), linguistiques ou culturels et leurs exploitations à travers les vastes réseaux aujourd'hui disponibles et de plus en plus accessibles par chacun de nous. Pour l'avenir, l'un des domaines les plus importants concerne sans aucun doute les modélisations sémantiques et les procédures d'accès au contenu telles qu'elles émergent aujourd'hui dans ce que l'on qualifie de Web sémantique et qui contribuent à l'amélioration des accès aux informations scientifiques, techniques et culturelles le plus souvent sous formes textuelles.

Nous disposons en France d'équipes de recherche de qualité en linguistique de corpus et en TAL et, s'il est vrai qu'un des objectifs premier de la linguistique reste de rechercher

des invariants sur le fonctionnement des diverses langues, il est nécessaire, pour permettre au français de rester présent dans ce champ fortement internationalisé, de poursuivre le développement de recherches, outils et ressources liés plus spécifiquement à notre langue. Il apparaît donc naturel et nécessaire de développer en France des études plus spécifiques sur la langue française, pour elle-même et en interaction avec des langues partenaires, mais aussi pour permettre au français de rester présent, à côté de la langue dominante qu'est aujourd'hui l'anglais, dans le plus grand nombre de champs applicatifs de notre société de l'information (gestion d'informations et accès par le contenu, traduction assistée par ordinateur, industries de la langue, etc.).

Il s'agit donc d'un enjeu considérable, sur lequel il convient de travailler de façon collective en s'appuyant les dispositifs existants. En particulier, la question de l'élaboration et la mise à disposition de ressources doit être faite en concertation avec les centres des ressources numériques, en particulier le CRNTL (<http://www.atilf.fr/>) et le CRDO (<http://www.lpl.univ-aix.fr/~crdo/>).

3 – CONCLUSIONS

La linguistique produit des savoirs qui ont différentes finalités comme :

- accroître la connaissance sur les langues et le langage ;
- satisfaire la demande accrue de connaissances linguistiques que réclame le développement de l'ingénierie linguistique ;
- répondre aux besoins de politique linguistique ;
- enrichir l'enseignement des langues ;
- contribuer au traitement de certaines pathologies concernant les atteintes de la voix, de la parole, du langage, de la perception auditive.

Soulignons à nouveau que les sciences du langage occupent une place unique vis-à-vis de la majorité des sciences humaines dans la mesure où à la fois le matériau empirique de ces dernières et les objets qu'elles produisent sont des discours, des productions langagières. De ce point de vue les sciences du langage – et en particulier les sous-disciplines qui étudient les textes, les discours, les conversations – peuvent apporter aux autres sciences humaines une aide pour l'analyse des données qu'elles ont recueillies même si les faits qui les intéressent ne sont pas de nature linguistique.

Par ailleurs, parce qu'ils sont aussi des acteurs sociaux, les linguistes ont également pour mission de mener une réflexion sur leur propre place et leur propre rôle qu'ils occupent dans la société et la cité ; la langue, rappelons le, reste un aspect important de l'identité d'une société, qui renferme de multiples enjeux.

Dans la mesure où il est dans l'esprit d'un rapport de ce type de s'interroger sur l'ancrage institutionnel et social de la discipline, il semble utile de :

- s'interroger sur la perception de la linguistique comme science afin de déconstruire une image erronée assimilée à l'étymologie, à la comparaison des langues, etc. qui perdure ;

- rendre plus lisible le rapport du linguistique au social, en insistant davantage sur la politique de la langue, la planification linguistique, le traitement des langues minoritaires, le traitement des langues (question de l'orthographe) ;

- réfléchir sur la perception du français (en France et à l'étranger), de la Francophonie ;

- travailler au développement de synergies entre le secteur des Lettres et les Sciences du

langage, en renforçant par exemple la tradition logicienne dans les cursus de Lettres comme cela se fait dans d'autres pays ;

- contribuer à une plus grande visibilité de l'enseignement en linguistique en construisant notamment un cursus commun.

ANNEXE

ANNEXE 1 : LISTE DES ACRONYMES

La première annexe est consacrée à la signification des sigles.

ILF	Institut de Linguistique Française (FR2393)
IST	Information Scientifique et Technique
OHLL	Origine de l'Homme, du Langage et des Langues
MPI	Max Planck Institute for Psycholinguistics
SDL	Sciences du Langage
TAL	Traitement Automatique des Langues
TUL	fédération de recherche en Typologie et Universaux Linguistiques (FR2559)
SOAS	School of Oriental and African Studies

