

La langue : entre usage quotidien et objet de connaissance

Décembre 2013

Claire Beyssade

On pourrait être tenté d'opposer, en linguistique comme dans l'ensemble des sciences humaines et sociales, peut-être même comme dans l'ensemble des sciences en général, deux types de connaissances, la connaissance savante d'une part et la connaissance vulgaire de l'autre. Les deux adjectifs ayant des connotations opposées, l'un de ces savoirs est valorisé par rapport à l'autre. Le savoir savant serait celui des spécialistes ou des professionnels d'une discipline, constituée en tant que telle. Le savoir vulgaire serait celui du commun des mortels ou de l'amateur. Dans le domaine de la physique par exemple, on pourrait dire que le savoir vulgaire correspond à la physique naïve, dont tout un chacun fait usage dans sa vie quotidienne quand il se déplace, qu'il doit traverser une rue et estimer la vitesse des voitures qui arrivent. L'opposition savant/vulgaire s'articule alors autour de deux pôles, au moins : (i) un savoir savant, produit par une communauté restreinte de spécialistes *vs.* un savoir vulgaire partagé par le plus grand nombre, (ii) un savoir savant élaboré selon certains critères qui en garantissent la scientificité *vs.* un savoir vulgaire élaboré pour et par une pratique.

Pour illustrer cette opposition entre vulgaire et savant dans le domaine des langues, on pourrait mettre en regard les règles des grammaires scolaires, comme celle qui stipule que tous les mots se terminant en *-ou* prennent un *s* au pluriel, à l'exception de *bijou*, *caillou*, *chou*, *genou*, *hibou*, *joujou* et *pou* et les travaux des linguistes qui mettent en lumière une régularité derrière ces exceptions. Nina Catach par exemple a montré que les mots en *-ou* qui prennent un *x* au pluriel sont ceux qui finissaient anciennement par un *l* mouillé, comme *pou* ou *genou* (en ancien français *pouil* et *genouil*). Pour *chou*, c'est la présence ancienne du *l* final (*choul* jusqu'en 1606) qui a favorisé le maintien du *x* au pluriel, comme dans *ciel* qui a pour pluriel *cieux*. Il y a donc bien d'un côté un savoir-faire et « des recettes » pour l'acquérir, et de l'autre une connaissance plus profonde des phénomènes, appuyée sur des explications : à une liste d'exception, le savant a substitué une règle générale qui permet de comprendre l'état actuel de la langue en recourant à la connaissance d'états antérieurs de celle-ci. Cette règle générale fait appel à des considérations diachroniques, inaccessibles à tous ceux qui maîtrisent parfaitement la langue mais en ignorent l'histoire. On pourrait aussi, pour illustrer la connaissance vulgaire, faire référence aux nombreuses idées reçues sur les langues, ces idées partagées par beaucoup, comme par exemple que certaines langues sont plus logiques ou plus faciles à apprendre que d'autres, idées qui donnent lieu régulièrement à la parution d'ouvrages grand public comme le *Catalogue des idées reçues sur la langue* de Marina Yaguello (1988) ou le livre de Chantal Rittaud-Hutinet (2011) *Parlez-vous français ? Idées reçues sur la langue française*.

Mais opposer ainsi de façon binaire deux types de connaissance reflète mal la réalité. En effet, quand bien même on considérerait qu'il y a aux deux bouts du spectre de la connaissance, la connaissance vulgaire et la connaissance savante, ce qui est intéressant, c'est d'étudier le continuum qui va d'une extrémité à l'autre, ainsi que les processus qui permettent de passer d'un type de connaissance à l'autre. En d'autres termes, les processus de vulgarisation et de savantisation des connaissances. Dans cet article, nous allons, en nous appuyant sur quelques exemples, essayer de montrer qu'il n'y a pas de frontière étanche entre connaissance vulgaire et savante et qu'il serait réducteur de voir dans la connaissance vulgaire soit des idées fausses, soit une simplification de la connaissance savante. On montrera, au delà

de la porosité¹ qui existe entre les deux types de connaissances, que, de même que les savoirs savants peuvent changer les représentations ordinaires, de même les savoirs vulgaires peuvent contribuer au renouvellement des savoirs savants. Il arrive en effet que sous la pression d'une demande venant du public ou orientée vers la maîtrise d'un savoir pratique par le plus grand nombre, la connaissance savante évolue. On en donnera deux exemples dans la première partie consacrée d'une part aux dictionnaires et à l'évolution des pratiques lexicographiques et d'autre part à la question de l'origine des langues et du langage. Dans la seconde partie, on s'interrogera sur les processus de légitimation du savoir savant et on verra, au travers de l'histoire des réformes de l'orthographe, le rôle que jouent les institutions et les groupes d'experts dans l'établissement de la connaissance savante. Dans une troisième partie, on précisera le parallèle qui peut être établi entre connaissance savante et connaissance vulgaire d'une part et connaissance théorique et connaissance pratique d'autre part. Ce point est particulièrement important en ce qui concerne les connaissances sur les langues, parce que tout le monde utilise quotidiennement le langage et de ce fait a le sentiment de connaître sa langue. Les linguistes ont donc dû réfléchir à la place de l'introspection dans leur discipline et à la façon d'objectiver les données linguistiques. Enfin, on terminera par quelques réflexions sur les différences entre connaissance vulgaire et idées reçues.

1) QUAND LE VULGAIRE INTERROGE LE SAVANT

La question du rapport entre les connaissances vulgaires et les connaissances savantes se pose de façon spécifique en linguistique, parce qu'il n'est pas besoin d'être linguiste pour parler une langue. En effet, chaque individu naît avec la capacité de parler une langue, développe cette capacité au cours du temps et devient locuteur natif de telle ou telle langue. Il en fait alors un usage quotidien et est conduit à réfléchir sur sa langue, tant au cours de sa scolarité que dans la vie de tous les jours, les occasions ne manquant pas de se demander si on s'est bien fait comprendre, si on a choisi le mot juste... Chacun peut donc prétendre connaître une langue (au moins), puisqu'il la parle couramment. Cela conduit beaucoup de gens à s'interroger sur ce qui fait la spécificité du linguiste : quelle compétence supplémentaire a-t-il ? Qui plus est, le linguiste est souvent perçu comme un littéraire et a du mal à se faire reconnaître comme un scientifique. On le confond régulièrement soit avec le grammairien (versé dans l'art de parler et d'écrire correctement), soit avec un spécialiste maîtrisant une ou plusieurs langues étrangères, en plus de sa langue maternelle. Pourtant la connaissance que le linguiste a des langues et du langage va bien au delà de la simple pratique d'une ou de plusieurs langues : il essaie de comprendre à la fois le fonctionnement propre des diverses langues du monde, mais aussi ce qu'il y a d'universel dans les différents systèmes linguistiques, de manière à décrire la grammaire universelle et à comprendre la faculté de langage.

Une seconde spécificité des linguistes est qu'ils utilisent le langage pour étudier le langage. Cela pourrait contribuer à brouiller la frontière entre savant et vulgaire à propos des langues. Mais en réalité, pour décrire les langues, les linguistes ont souvent recours à une terminologie spécifique (faite de mots techniques comme *conjonction*, *syntagme*...) utilisée exclusivement pour parler du langage. Plus la discipline s'est développée, plus ce métalangage s'est enrichi. Il est devenu tellement spécialisé qu'il n'est quelquefois plus du tout accessible aux non linguistes : un fossé peut séparer le vocabulaire de la connaissance savante et le vocabulaire de la connaissance vulgaire sur les langues. Mais ce n'est pas toujours le cas. On le voit avec les dictionnaires d'usage, qui s'interdisent l'emploi de termes techniques dans leurs définitions. L'histoire des dictionnaires permet aussi de montrer

¹ Nous reprenons ici le terme utilisé par C. Prochasson dans ce même volume.

comment une discipline nouvelle, la lexicographie, s'est peu à peu constituée par un retour de la connaissance vulgaire sur la connaissance savante, les éditeurs ayant vu dans la production de dictionnaires à destination du grand public un nouveau marché.

1.1 La lexicographie

Bien qu'on dise couramment « chercher dans le dictionnaire », il n'existe pas un dictionnaire mais une multiplicité de dictionnaires, qui se distinguent par la taille (qui va du dictionnaire de poche au dictionnaire encyclopédique en plusieurs volumes), le type de descriptions qu'ils s'autorisent (dictionnaire général, dictionnaire des synonymes...), le champ lexical choisi (dictionnaire de la médecine, de l'argot...), le nombre de langues décrites (dictionnaires unilingues ou bilingues) mais aussi le public qu'ils visent. Comme le dit Cerquiglini (1992), « le dictionnaire résume, concentre et représente, aux yeux de beaucoup, la langue elle-même, avec laquelle il a une relation essentielle, bien que jamais achevée. Monolingue, bilingue ou plurilingue, classique ou plus moderne, général ou spécialisé, de l'humble glossaire terminologique dont le but est précis à la fascinante encyclopédie qui dit "tout sur tout", c'est un monde de dictionnaires qui encadre et organise notre vocabulaire et notre réflexion ». On peut penser que la variété de ces dictionnaires donne une bonne idée de la multiplicité des formes de savoirs qu'on trouve dans ce domaine : les dictionnaires spécialisés enregistreraient une connaissance savante à la différence des dictionnaires d'usage, à destination du grand public, qui enregistreraient une connaissance vulgaire ; de la même façon, la taille du dictionnaire reflèterait directement la nature plus ou moins savante de la connaissance qui s'y trouve. Mais les choses sont en réalité plus complexes et un rapide parcours de l'histoire des dictionnaires et des pratiques lexicographiques (repris à Pruvost (2002)) permet de mieux voir comment s'articulent, dans des productions effectives, connaissances communes et connaissances savantes.

Pour commencer, précisons qu'on ne va considérer ici que les dictionnaires unilingues consacrés au français. Ce type de dictionnaires est apparu au XVII^{ème} siècle. Auparavant n'existaient que des dictionnaires plurilingues, mais comme le dit Pruvost, c'est dans les répertoires plurilingues que prennent souche les dictionnaires unilingues. En 1531, Robert Estienne publie un dictionnaire latin-français, dans lequel chaque mot latin est suivi de son équivalent français et parfois de quelques explications en français. En 1539, l'imprimeur propose de placer les mots français en première position et publie *le Dictionnaire françois-latin contenant les motz et manieres de parler françois tournez en latin*. Il ne reste plus qu'à enlever les mots latins pour avoir un dictionnaire unilingue. Le premier dictionnaire français-français date de 1680 : c'est le *Dictionnaire français contenant les mots et les choses* de Richelet. Il décrit le bel usage des mots, au moyen d'exemples choisis dans l'œuvre d'auteurs comme Boileau, Molière, Pascal, Vaugelas. A peu près à la même époque paraît le *Dictionnaire Universel, contenant généralement tous les mots françois tant vieux que modernes et les termes de toutes les Sciences et des Arts* de Furetière. A la différence du précédent, il ne contient pas de citations d'auteurs, mais s'attache aux idées et aux choses décrites par les mots plus qu'à l'usage du mot dans la langue. En ce sens, il s'apparente à un dictionnaire encyclopédique. Enfin en 1694 paraît la première édition du *Dictionnaire de l'Académie française*, résultat d'un travail commencé en 1635 sur la demande de Richelieu. Dans cette première édition du dictionnaire de l'Académie, les mots n'étaient pas présentés par ordre alphabétique, mais regroupés par racines, ce qui explique en partie le mauvais accueil qu'il reçut du public, qui le trouva trop « savant ». A titre d'exemple, il fallait chercher les mots *dette*, *débiter* et *redevance* sous l'entrée *devoir*. En revanche, le choix fait par l'académie de décrire l'usage contemporain en construisant des exemples plutôt qu'en s'appuyant sur des citations d'auteurs était très moderne et novateur.

Le XVIII^{ème} siècle est marqué par la publication de *l'Encyclopédie ou le Dictionnaire raisonné des sciences, des arts et des métiers* de Diderot et D'Alembert : trente-cinq volumes dont dix-sept de textes, plus des suppléments, des tables analytiques et des planches. Il ne s'agit donc pas d'un dictionnaire de langue mais d'un dictionnaire encyclopédique dont l'objectif est de présenter de façon synthétique la connaissance de l'époque. Les entrées y sont classées par thèmes.

On distingue ensuite deux périodes dans le XIX^{ème} siècle. Au cours de la première moitié du siècle parurent des ouvrages comme *Le Dictionnaire universel de la langue française* de Boiste (1800), le dictionnaire de Napoléon Landais (1834) ou celui de Bescherelle (1845). Tous se caractérisent par une augmentation très importante du nombre des entrées lexicales et un désintérêt évident pour les définitions et les principes lexicographiques. Quant à la seconde moitié du XIX^{ème} siècle, elle se caractérise par l'essor de la linguistique historique, le développement des recherches étymologiques et l'établissement des règles de phonétique historique. Émile Littré publie en 1872 le *Dictionnaire de la langue française* en quatre volumes, qui fait une large part à l'histoire des mots, à l'étymologie et à la filiation historique des sens d'un mot tout en s'appuyant sur des citations d'auteurs. Cependant, ces citations ne sont jamais postérieures à 1830. En 1856 Pierre Larousse publie le *Nouveau dictionnaire de la langue française*, dictionnaire de petite taille et destiné à un public scolaire. Il connaît un succès considérable. Ce sera l'ancêtre lointain du Petit Larousse illustré dont la première édition est de 1906 et que l'on doit à ses successeurs. Après avoir publié ce petit dictionnaire, Pierre Larousse se lance dans un projet d'une tout autre envergure : le *Grand dictionnaire universel du XIX^{ème} siècle* en quinze volumes, qu'il publie en 1876. En 1904 paraît une version amincie de ce grand dictionnaire : c'est le *Nouveau Larousse illustré* qui ne compte plus que sept volumes.

Au XX^{ème} siècle, les dictionnaires de petite taille, en un ou deux volumes, se multiplient sans pour autant éclipser les dictionnaires encyclopédiques. Deux éditeurs se partagent la production : Larousse et Robert.

En 1910 paraît le *Larousse pour tous* en deux volumes, rebaptisé *Larousse Universel* en 1923, et *Nouveau Larousse Universel* en 1948. Ces deux dictionnaires paraissent sous la direction de Jean Dubois : d'une part, en 1963, le *Grand Larousse Encyclopédique* en dix volumes, auquel 700 spécialistes ont participé ; d'autre part en 1967 le *Dictionnaire français contemporain*, un ouvrage en un volume, de format très réduit et destiné au public scolaire.

A la même période, Paul Robert achève le *Dictionnaire alphabétique et analogique de la langue française* et propose ainsi en 1964 un ouvrage de six volumes qui rappelle le Littré avec ses citations extraites de textes littéraires plus récents. Et en 1967 paraît le *Petit Robert*, comparable par la taille au *Petit Larousse illustré*, mais différent puisqu'il s'agit d'un dictionnaire de langue et non plus d'un dictionnaire encyclopédique. Dans les années 90, les ouvrages en un ou deux volumes se multiplient : dictionnaires pour enfants, dictionnaires pour collégiens, dictionnaires de noms propres...

Parallèlement à ces productions éditoriales, un grand projet s'est développé de 1971 à 1994 au sein du CNRS : l'élaboration d'un dictionnaire de la langue française des XIX^{ème} et XX^{ème} siècles destiné à remplacer le Littré, le *Trésor de la langue française*. Il présente plus de 100 000 mots et leur histoire, 270 000 définitions et 430 000 exemples. En tout 16 volumes publiés en format papier en 1994, puis sous forme électronique en 2004. *Le Trésor de la langue française informatisé* est en accès libre sur le site de l'ATILF et permet à la fois une visualisation article par article et l'accès aux données via des requêtes spécifiques : on peut par exemple rechercher les mots d'origine anglaise, les mots associés à un domaine particulier ou les mots caractéristiques d'une région (comme le français du Québec).

Que nous montre ce rapide aperçu de l'histoire des dictionnaires ? Qu'il y a une continuité des formes de savoir et qu'il est impossible de ne considérer que deux termes qui seraient connaissance savante vs. connaissance vulgaire. Que la vulgarisation n'est pas identifiable à une réduction des connaissances, pas plus que la savantisation n'est liée à une augmentation quantitative des informations recensées, et enfin qu'il peut y avoir un retour de la connaissance vulgaire sur la connaissance savante. En effet, les processus qui ont accompagné la mise en place de dictionnaires d'usage ont donné lieu à l'émergence d'une discipline, la lexicographie, qui se distingue à la fois de la lexicologie, de la sémantique lexicale et de l'étymologie.

La variété des dictionnaires donne une illustration de la difficulté qu'il y a à définir la connaissance savante et la connaissance vulgaire et de la continuité qui existe entre ces deux formes de savoirs. Il existe des types de dictionnaires très différents et les noms qu'ils prennent laissent à penser qu'ils sont plus ou moins savants ou du moins qu'ils s'adressent à un public plus ou moins savant : *mon premier dictionnaire*, *le dictionnaire des écoliers*. Mais sur quels critères s'appuyer pour dire qu'un dictionnaire est plus ou moins savant ? Sans doute pas la taille du dictionnaire. Au XIX^{ème} siècle par exemple, les dictionnaires ont multiplié les entrées lexicales et sont devenus de plus en plus gros. Mais comme cette augmentation ne s'est accompagnée d'aucune réflexion sur la façon d'organiser ces entrées, il n'y a pas lieu de considérer qu'ils étaient plus savants que les dictionnaires précédents. Au contraire, Pruvost résume la période en disant que « les titres [de ces dictionnaires] sont révélateurs d'une surenchère qui se situe davantage sur la quantité que sur la qualité ». Par ailleurs, il existe de petits dictionnaires très spécialisés, comme le *Dictionnaire du marketing* d'Albertini et al. (2008) ou le *Dictionnaire de la mythologie grecque et romaine* de P. Grimal (1996). Ils sont petits parce qu'ils portent sur un champ lexical restreint, souvent technique mais ils s'adressent clairement à un public de spécialistes et en ce sens enregistrent une connaissance savante.

De façon générale donc, il n'y a pas de raison de penser que la connaissance savante est celle qui occupe le plus de volume, ni que la connaissance vulgaire s'obtient par réduction de la connaissance savante. On a vu que certains auteurs ont écrit à la fois des dictionnaires de petite et de grande tailles. Jean Dubois par exemple a d'abord dirigé l'édition du *Grand Larousse Encyclopédique*, puis celle du *Dictionnaire du Français Courant*. Il est certain que pour passer de dix volumes à un seul, il n'a pas pu se contenter d'éliminer certains articles et de demander aux rédacteurs des autres de réduire leur texte. Le *Dictionnaire du Français Courant* a été conçu d'une façon tout à fait originale, Dubois ayant pris le parti de se focaliser sur le français en synchronie, et de dégrouper systématiquement les entrées correspondant à des sens différents (plusieurs articles au lieu d'un seul sont présentés par exemple pour le mot "classe"). A l'époque, ces choix « avaient fait l'effet d'une révolution lexicographique », selon Pruvost (2002). On voit donc que ce qui fait la spécificité d'un dictionnaire, c'est au moins autant la façon dont sont rédigées les définitions que le nombre des entrées. Le rôle du lexicographe n'est pas seulement de répertorier les mots d'une langue, mais bien de les classer et de les définir. L'histoire des dictionnaires montre une évolution dans les principes de classement et dans l'élaboration des définitions. Ce travail s'est systématisé au cours du temps. Le lexicographe a cherché des principes réglant l'écriture de chaque article, il a réfléchi aux primitives à utiliser pour classer les mots en catégories, a essayé de hiérarchiser et de systématiser l'information à placer dans chaque article, de manière à éviter plusieurs écueils : la dispersion des informations, la redondance, la circularité des définitions, le manque de cohérence des articles entre eux. On peut rappeler que c'est faute d'avoir choisi un principe simple de classement des entrées que le premier dictionnaire de l'académie n'a pas rencontré le succès escompté. De même le choix fait par Littré d'ordonner les différents sens d'un même mot selon la date d'apparition de ce sens dans la langue, qui le conduit à

distinguer jusqu'à 40 « sens » pour un même mot, a été jugé par beaucoup comme problématique, d'autant qu'il ne reflète en rien la réalité de l'usage. La conception des articles est donc un trait essentiel du dictionnaire et c'est sans aucun doute dans l'organisation des articles et les principes qui la sous-tendent que se situe le caractère plus ou moins savant de l'ouvrage.

L'histoire des dictionnaires nous montre qu'on peut distinguer deux façons de concevoir les articles de dictionnaires : en érudit, ou en privilégiant l'idée de système. Josette Rey-Debove a très bien décrit le dilemme devant lequel tout lexicographe se trouve. C'est que dans les années 70, alors que le structuralisme pesait sur toute la pensée de l'époque, ce dilemme se faisait sentir encore plus lourdement qu'à tout autre période. Elle écrivait alors : « La lexicographie vit actuellement une crise. Au moment même où les progrès et la vulgarisation des études diachroniques allaient servir de base à une meilleure description lexicographique, l'avènement du structuralisme synchronique remet en question les fondements même de la description, sans proposer de modèles suffisants. Le lexicographe dont le propos était une amélioration constante du dictionnaire par l'utilisation de données toujours plus exactes et plus précises, se trouve dans l'obligation de repenser son système métalinguistique en théoricien et non en érudit. » Rey-Debove (1971, p. 316). Il existe donc selon elle une tension entre le théoricien et l'érudit : le rôle du lexicographe est-il d'enregistrer le maximum d'informations dans le dictionnaire, ou de mettre en avant les seules informations pertinentes, quitte à écarter ou à passer sous silence ce qui ne trouve pas de place dans le système général des oppositions de la langue ? Si l'on connaît les dangers d'une position trop fortement structuraliste, on doit aussi souligner les dangers de l'érudition. L'érudit est certes savant, mais, sous prétexte de restituer des connaissances historiques sur un mot et son évolution dans le temps, il faut bien qu'il se garde de reconstruire une histoire, une narration empreinte de fiction. Nombreux sont ceux qui ont imaginé l'histoire d'un mot ou d'une règle. Or il y a une différence de nature entre le discours scientifique et la narration. Le plus savant est aussi quelquefois le moins scientifique². C'est ce que ne comprend pas toujours l'étymologiste habité par les fantasmes de la reconstruction des origines et à qui il manque quelquefois l'idée de système.

Donc, s'il est toujours intéressant d'accumuler des connaissances historiques sur la langue, l'érudition peut néanmoins présenter un danger. Par ailleurs, la vulgarisation ne se fait par simple réduction des données de la connaissance savante. Ce que nous montre l'histoire des dictionnaires, c'est un mouvement inverse, qui va de la connaissance vulgaire vers la connaissance savante. Les innovations, en matière de lexicographie, sont venues des petits dictionnaires à destination du grand public. Pour écrire des dictionnaires généraux à destination du plus grand nombre, le lexicographe a dû profondément repensé l'organisation et l'écriture des articles. Ces petits dictionnaires sont en général plus « raisonnés » que les dictionnaires encyclopédiques plus volumineux. On peut même dire que la lexicographie en tant que discipline appliquée, ayant pour objet l'élaboration de dictionnaires, est née dans les maisons d'édition et qu'en réfléchissant aux techniques de confection des dictionnaires, les lexicographes ont obligé les lexicologues, qui appréhendent de façon théorique l'étude générale du lexique, à préciser des notions fondamentales comme celle de mots, de polysémie et d'ambiguïté. Même si le lexicographe se veut un chercheur seulement préoccupé par la rigueur scientifique de son analyse et de sa description lexicale, sans se préoccuper des contraintes inhérentes à l'élaboration et à la diffusion du dictionnaire en tant que produit vendu, il est confronté à des difficultés pratiques qui le conduisent à poser des questions nouvelles aux lexicologues. La production des dictionnaires d'usage, qui peuvent être vus comme des objets de connaissance commune dans la mesure où ils sont produits pour être mis

² On observera quelque chose de comparable avec les réflexions sur l'orthographe.

à la disposition du plus grand nombre, a donc conduit une discipline à se constituer en fondant sa pratique sur des principes scientifiques : définitions des primitives lexicographiques (catégories) et élaboration de règles lexicographiques permettant d'assurer à la fois la cohérence du dictionnaire, sa compacité et sa complétude.

On retiendra donc de cet exemple que, de même que la vulgarisation (au sens de production de connaissance vulgaire) doit éviter l'écueil de la simplification, la connaissance savante doit éviter l'écueil de l'érudition ; mais aussi que la production d'objets à destination du grand public (les dictionnaires courants) a conduit à un changement des pratiques et à l'émergence d'une discipline, la lexicographie.

1.2 L'origine des langues : une demande du vulgaire au savant

La question de l'origine des langues suscite depuis longtemps l'intérêt des hommes : quand et comment les langues sont-elles apparues, y avait-il une ou plusieurs langues originelles, et à quoi ressemblaient-elles ? Mais pendant longtemps, ces questions n'ont pas pu être abordées de façon scientifique et cela a conduit certaines sociétés savantes comme la Société de Linguistique de Paris à indiquer explicitement dans son règlement, en 1866, qu'elle ne recevrait aucune communication sur ce sujet. Ce n'est qu'assez récemment que cette mention a disparu. Depuis une vingtaine d'années, la question de l'origine des langues a connu un regain d'intérêt, les recherches sont devenues pluridisciplinaires et elles mettent en jeu à la fois les savoirs des paléanthropologues et ceux des linguistes spécialistes des proto-langues. On distingue désormais deux aspects dans la question : d'une part, quelle est l'origine du langage (ou en d'autres termes, quand l'homme a-t-il commencé à parler) et d'autre part, quelle est l'origine des langues (les différentes langues actuelles ont-elles une racine unique ?). On va retracer brièvement l'histoire de la question de l'origine des langues³, marquée un temps par un conflit entre connaissance savante et idéologie, puisque la question a été mise à l'index. Aujourd'hui, sous la pression de la demande du grand public, la question est de nouveau posée dans la recherche contemporaine et elle suscite beaucoup de débats concernant à la fois la pertinence de la recherche, l'évaluation des programmes mis en place et l'évaluation de leurs résultats. Selon Auroux (2007), ces débats conduisent à s'interroger sur les liens entre connaissances savantes et sciences ainsi que sur les dérives du paradigme naturaliste dans les sciences humaines.

Auroux rappelle qu'on trouve dans le mythe biblique une réponse à la question de l'origine des langues : il y a « d'abord la donation des noms aux animaux par Adam à l'invitation de Dieu (*Genèse* 2, 19-20) ; puis l'épisode du déluge et l'installation de la dynastie de Noé et ses fils (*Genèse* 8) ; enfin [...] l'épisode de Babel (*Genèse* 11). [...] La cause de la naissance du langage (acte d'Adam décidé par Dieu) n'est pas la même que celle de la différenciation des langues (intervention externe destinée à rompre l'intercommunication » (Auroux, 2007, p. 20). Peu après le déluge, les hommes atteignent une plaine, s'y installent et entreprennent de bâtir une ville et une tour dont le sommet touchera le ciel. Dieu estime que s'ils y arrivent, ils seront trop puissants car plus rien ne leur sera inaccessible. Il décide alors de brouiller leur langue afin qu'ils ne se comprennent plus et il les disperse sur toute la surface de la terre. « Descendons ici même, confondons leur langage en sorte qu'ils ne comprennent plus le langage les uns des autres » est-il écrit dans le livre 11 de la *Genèse*. A cette explication mythologique ont succédé d'autres suggestions et on peut distinguer quatre moments importants qui marquent des tournants dans l'histoire de la question⁴.

³ Cette présentation reprend les grandes lignes du texte de Sylvain Auroux (2007).

⁴ Pour une présentation plus approfondie de la question, le lecteur pourra se reporter à Auroux (2007), qui parle de cinq articulations ponctuant la question. Aux quatre évoquées dans notre

Il y a d'abord eu une laïcisation de la question à l'époque des Lumières. En 1769, l'Académie de Berlin met au concours la question suivante : « En supposant que les hommes soient abandonnés à leurs facultés naturelles, sont-ils en état d'inventer le langage ? Et par quels moyens parviendront-ils à cette invention ? ». Il s'agit donc de chercher une explication non plus divine, mais rationnelle, faisant appel à des processus naturels, ordinaires et stables. La laïcisation de la question ouvre deux voies : (i) la construction de modèles théoriques d'émergence du langage (comme celui de Condillac, qui dit clairement que l'arbitraire du signe, c'est le principe théorique à l'origine de l'histoire et de la culture, que le passage du signe naturel au signe arbitraire inaugure la possibilité pour le sujet humain de manipuler les signes à son gré et d'entrer dans le monde du langage) ; (ii) une approche historique et philologique.

Aux XVIII^{ème} et XIX^{ème} siècles, la discipline prend une orientation naturaliste (une langue est une espèce naturelle qui naît, croît et meurt) et les linguistes pensent faire avancer la question grâce aux méthodes de la grammaire comparée. Ils mettent au point des méthodes codifiées pour comparer les langues. De manière générale, quand les langues sont bien connues, ils mettent en parallèle des dictionnaires et des éléments grammaticaux ; quand elles le sont moins, ils utilisent des tableaux de mots élémentaires ou la traduction d'un texte court. Si la seconde voie permet d'éclatants succès quand on est à l'intérieur d'un groupe de langues apparentées, elle peut mener à des conclusions erronées lorsqu'on l'étend sans restriction (le programme de Court de Gébelin, au XVIII^{ème} siècle, en est un exemple). Pour un comparatiste qui cherche à regrouper les langues par familles, le problème essentiel n'est pas l'extension des comparaisons, mais leur limitation : avec cette méthode, il est très difficile, voire impossible, de prouver que deux langues ne sont pas parentes, parce que tout rapprochement est irréfutable. Cependant, comme le montre Auroux (2007, p. 40), les travaux de Rask, Grimm et Bopp en grammaire comparée viennent mettre un terme à ce travers et constituent une révolution épistémologique. Ils opèrent un changement de programme scientifique, en considérant que l'unité de référence n'est plus le mot mais le son (le phonème correspond ici à une lettre), qui change toutes ses occurrences en fonction de lois phonétiques à une époque et pour des langues données. Les changements deviennent spécifiques à chaque langue.

De façon étonnante, en 1866, alors que la méthode comparatiste est en pleine extension, la Société de Linguistique de Paris se crée en inscrivant dans ses statuts qu'elle n'admettra « aucune communication concernant soit l'origine du langage, soit la création d'une langue universelle ». Henri, un grammairien de l'époque, résume bien le point : « Que le linguiste doive s'interdire toute recherche sur l'origine du langage, c'est un point qui semble définitivement acquis, tout au moins parmi les linguistes, si paradoxale qu'en soit la première apparence : l'origine du langage n'est pas, *a priori*, un problème linguistique, puisque la linguistique ne se propose pour objets que des langues toutes formées, dans leur état actuel, historique ou préhistorique et qu'il ne lui est donné que de constater l'évolution, jamais la naissance d'un langage. »

Ce malaise des linguistes face à la question des origines ne disparaît véritablement et officiellement qu'en 1975, quand une conférence est organisée sous les auspices de l'Académie des sciences de New-York avec pour thème : *Origins and Evolution of Language and Speech*. C'est cette conférence qui marque le grand retour de la question dans les programmes de recherche et le renouveau de la question dans le contexte d'une recherche

texte, il en ajoute une cinquième, correspondant aux discussions philosophiques qui ont conduit dans les années 1960 et 1970, dans la communauté francophone, à distinguer entre *commencement* et *origine* et à mettre en question l'idée d'une causalité entre des faits relevant de l'histoire biologique ou sociale de l'humanité et la structure du langage (cf Ibid. p. 24 et sq.)

dominée par le paradigme de la naturalisation des sciences humaines. Rappelons que Chomsky soutient que la linguistique est une branche de la biologie humaine, mais aussi que la découverte et la diffusion des théories évolutionnistes a permis de réalimenter le débat : l'évolutionnisme interfère avec le débat sur l'origine puisqu'en présentant l'origine animale de l'homme, il induit logiquement l'idée d'une émergence de la faculté de langage. Auroux (2007, p. 95) montre que ce renouveau des études sur l'origine des langues et du langage dans la seconde moitié du XX^{ème} siècle s'appuie sur (i) des avancées de la paléontologie humaine, (ii) l'extension du paradigme généalogique à la préhistoire par Swadesh, Greenberg et Ruhlen⁵ notamment, (iii) la comparaison entre la classification génétique des populations et la classification des langues (développée notamment par Cavalli-Sforza) et (iv) l'idée d'une nouvelle synthèse disciplinaire, menée dans le cadre d'une domination des chomskyens en matière de théorie linguistique, à un moment où se multiplient les études psycholinguistiques et celles sur les primates et leur capacité à communiquer.

Auroux souligne la fragilité des résultats qui servent de point de départ à ce renouveau du questionnement sur les origines. En ce qui concerne les avancées de la paléontologie humaine, il rappelle que si l'archéologie plaide en faveur d'une sorte de révolution technique et cognitive entre le paléolithique moyen et le paléolithique supérieur, en revanche le lien entre l'objet technique et le langage a été remis en cause après la découverte d'outils fabriqués chez les groupes de chimpanzés (voir Mc Grew, 1992).

A propos de l'extension du paradigme généalogique à la préhistoire développé par Swadesh, Greenberg ou Ruhlen, il note que la glottochronologie a été largement contestée et qu'on a montré que cette technique d'appareillage par le vocabulaire ne donnait pas de meilleurs résultats que des rapprochements aléatoires (voir Ringe, 1992). Il faut se souvenir que dans tous les cas où l'on disposait d'autres méthodes de datation, notamment, parce qu'il s'agissait de langues écrites disposant de documentation historique, on a pu vérifier que les datations données par la glottochronologie n'étaient pas validées. Autant dire qu'on a une réfutation empirique de cette méthode. Il rappelle aussi que l'étymologie globale prônée par Ruhlen refuse les principes fondateurs du comparatisme dont on a déjà mesuré l'importance dans l'histoire. Elle choisit le mot plutôt que le son comme unité de changement et en rejetant l'idée de lois, ne considère plus la langue comme un système. La réception des travaux de Ruhlen chez les linguistes a été assez dure. Les chercheurs de l'institut de la communication parlée de Grenoble ont monté un programme de simulation de la méthode utilisée. Les résultats sont impressionnants : dès que l'on utilise plus de deux changements sémantiques, la probabilité que les rapprochements soient dus au hasard avoisinent 1. Et Ruhlen a utilisé plus d'une dizaine de changements sémantiques. Un second problème vient s'ajouter au précédent : on peut partir des formes attestées et remonter à des proto-langues, mais ces proto-langues sont des entités abstraites, et les unités qu'on fait émerger, les racines, ne sont pas des mots. On ne peut en aucun cas reprendre la même méthode et comparer les proto-langues entre elles. La véritable question en matière d'origine du langage est le passage du signe au mot. On a tort de qualifier de mots les éléments des proto-langues. Par ailleurs, quand on franchit le seuil de la confirmation par les attestations, les comparatistes considèrent qu'on rentre dans le domaine de l'hypothétique (celui où il n'y a plus de fait pour confirmer ou infirmer) et que l'on cesse d'être dans la linguistique, et peut-être plus largement dans la science. Du coup, si on utilise des entités reconstruites pour effectuer des comparaisons et des regroupements en super-famille, on est dans le domaine de l'hypothétique, on a franchi la limite déjà posée par

⁵ Les linguistes américains ont tenté d'apporter une contribution proprement linguistique à la préhistoire du langage. Swadesh a inventé une méthode mathématique permettant de dater la séparation des idiomes (la glottochronologie).

Bopp concernant l'arbitraire des racines. Ces deux problèmes expliquent pourquoi, selon Auroux, certains ont qualifié l'étymologie globale de linguistique pré-copernicienne.

Quant au parallèle à établir entre langues et patrimoine génétique des populations, il est problématique parce qu'il prend pour point de départ une idée fautive, à savoir qu'il y aurait des familles de langues et une généalogie des langues comparable aux familles humaines et à la généalogie des espèces. Mais si on parle de familles de langues, si on dessine un arbre des langues, avec des ramifications issues de la proto-langue indo-européenne, c'est parce qu'on fait un usage métaphorique de cette notion. L'arbre est trompeur, non seulement parce qu'il donne l'illusion d'une généalogie, mais parce qu'il occulte les phénomènes de croisement, de substrat, les accidents culturels tels que la sélection de certains états de langue pour servir à des fonctions particulières, comme c'est le cas pour les langues religieuses et sacrées. L'arbre généalogique des langues n'est qu'une métaphore trompeuse, tout au plus nos classifications peuvent-elles identifier des lignées, c'est-à-dire des regroupements dans des ensembles distincts, mais cela n'exclut nullement d'autres formes de relations : typologie, convergence... Pourtant, les scientifiques qui travaillent aujourd'hui sur cette question continuent de s'appuyer sur cette idée : est-ce parce que, si elle n'est pas scientifique, elle peut néanmoins avoir une valeur heuristique ? Enfin, le concept de langue mère va de pair avec l'idée que le nombre des langues va croissant puisque la seule évolution possible est la division. Mais il faut compter avec les langues mortes. Quand bien même on pourrait remonter avec des méthodes assurées à une seule langue, qui unifierait toutes les langues existantes, cela ne prouverait pas qu'il n'y en ait pas eu d'autres disparues avec leur descendantes. D'autre part, même si on imagine que les données de la paléontologie nous amènent à admettre que le peuplement de la planète provient d'un groupe originaire, qu'est-ce qui nous empêche de penser que dans ce même groupe plusieurs langues ou dialectes se sont développés.

Tout cela montre combien les débats sur l'origine des langues, les programmes de recherche qui s'y consacrent et l'évaluation de leurs résultats sont controversés. Ils opposent de fait deux clans constitués. Il y a d'une part ceux qui soutiennent que cette recherche n'est pas scientifique ou ne s'oriente pas dans la bonne direction. C'est le cas d'Auroux qui conclut que pour le moment, on est face à des « spéculations sans possibilité de preuves empiriques » et que de façon générale « l'enthousiasme ne suffit pas à faire de la science ». Selon lui, « ou il y a des lois d'évolution linguistique (biologiques ou autres) et certains faits permettent d'inférer quelque chose sur l'origine des langues, ou les phénomènes linguistiques sont des entités historiques contingentes, et l'absence de faits datant de leur origine est indépassable ».⁶ D'autre part il y a ceux qui soutiennent qu'il faut continuer à chercher, en l'absence même de données nouvelles et de méthodes. Sur cette question la science n'en étant qu'à ses débuts, il faudrait la laisser tâtonner. Quoi qu'il en soit, même si les recherches qui sont poursuivies aujourd'hui ne permettent que d'élaborer des hypothèses et ne fournissent pas de certitude, elles cherchent à se conduire dans un cadre scientifique⁷. Et ce qui est certain, c'est que si de nouveaux programmes de recherche sont mis en place et financés, c'est sous la pression de la forte demande du grand public et en raison de la forte médiatisation de la question.

2) LA LEGITIMATION D'UN SAVOIR SAVANT : L'EXEMPLE DES REFORMES DE L'ORTHOGRAPHE

⁶ Auroux, 2007, p. 73 et sq.

⁷ C'est ce que montrent entre autres les deux ouvrages suivants : *Aux Origines des langues et du langage*, sous la direction de Jean-Marie Hombert, Fayard, 2005 ; Pascal Picq, Laurent Sagart, Ghislaine Dehaene, Cécile Lestienne, *La plus belle histoire du langage*, Seuil, Paris 2008.

On peut aussi s'interroger sur ce qui fonde un savoir savant comme savant. De ce point de vue, l'exemple de l'orthographe française et des réformes qui ponctuent son histoire est instructif. L'orthographe française est très compliquée, si on la compare à l'orthographe d'autres langues romanes comme l'espagnol ou l'italien par exemple. Cela vient de ce qu'elle n'a jamais été fondamentalement réformée. En effet, de nombreuses réformes ont été proposées, mais elles n'ont été adoptées que partiellement, et de façon régulière les débats sur l'orthographe ont cristallisé des clivages très forts dans la société. On pourrait penser que la solution aurait dû être de demander à un groupe d'experts de se prononcer sur les réformes à faire, mais le problème est de déterminer quels sont les experts en matière de langue et d'orthographe. Qui a la légitimité pour décréter que telle ou telle réforme doit être entreprise ? Les débats récents nous ont montré que plusieurs communautés revendiquent une telle compétence. Il y a d'un côté les linguistes qui dans leur grande majorité sont partisans d'une réforme, ainsi que les enseignants (notamment les instituteurs qui reconnaissent que le coût d'apprentissage de l'orthographe française, en temps et en efforts, est trop élevé pour des résultats médiocres). Dans l'autre camp on trouve des écrivains, des intellectuels, des utilisateurs de la langue qui voient dans toute réforme de l'orthographe un recul. Ils soutiennent entre autres choses que si l'orthographe française est complexe, c'est qu'elle reflète l'histoire de la langue et que choisir de réduire cette complexité, c'est effacer de la graphie des informations sur l'origine des mots. Force est de constater cependant que bien des exemples avancés pour défendre cette seconde position sont mal analysés, et il arrive que l'histoire associée à l'orthographe d'un mot soit erronée et s'apparente à de la fiction.

L'écriture est une pratique qui est normalisée. L'orthographe sert à assurer une intercompréhension des écrivains, en associant à chaque mot de la langue orale une (et le plus souvent une seule) graphie. L'orthographe d'une langue donnée peut être fondée sur des critères phonétiques ou phonologiques et dans ce cas la correspondance entre graphème et phonème est la règle : on associe à un mot une représentation de son image sonore. Il est évident qu'il n'y a pas de correspondance graphème-phonème parfaite et que puisque deux individus ne prononcent jamais de façon absolument identique le même mot, il faut bien qu'il y ait un écart entre ce qui est dit et ce qui est écrit. Mais l'orthographe phonétique ou phonologique se fixe comme principe de rester aussi proche que possible du son. A l'opposé, l'orthographe peut se fonder sur l'étymologie, et dans ce cas, l'orthographe garde dans la forme écrite des traces de l'histoire du mot, de son origine, de son évolution. On dit souvent que le français est une langue de ce type et en particulier que l'accent circonflexe sert à marquer l'existence d'un s dans la racine du mot considéré. Ainsi, si le nom « île » porte un accent circonflexe, c'est parce qu'il vient du mot latin *insula* dans lequel on observe la présence d'un s. Cependant, comme le montre Bernard Cerquiglini dans son livre « L'accent du souvenir » paru en 1995, il serait absolument faux de prétendre que l'accent circonflexe résulte toujours de la disparition d'un s, ou indique l'existence de dérivés du mot comportant un s. Tel est le cas pour *hôpital* qu'on peut rapprocher de *hospitalisation*, mais pas pour *âme* qui peut être rapproché de *animé* mais d'aucun mot comportant un s. En réalité, en retraçant l'histoire de l'accent circonflexe, de son apparition et de sa potentielle disparition, Cerquiglini donne un aperçu de l'histoire des réformes de l'orthographe, et cette réflexion conduit à poser la question de la légitimisation des savoirs savants.

Cerquiglini explique qu'on ne peut discuter de la légitimité de la disparition de l'accent circonflexe, sans préciser au préalable quand et pourquoi cet accent est apparu. Le français se distingue des autres langues romanes en ce qu'il a été soumis à une influence germanique entre le V^{ème} et le X^{ème} siècle au niveau phonétique. D'où une tendance à voir apparaître un fort accent d'intensité, accent qui a d'importantes conséquences sur l'évolution des voyelles et la transformation des mots : une voyelle accentuée, par exemple, prononcée avec une plus

grande force, a tendance à se diphtonguer. Il s'ensuit, selon Cerquiglini, que la langue parlée dans la Gaule du nord, à la différence des autres langues romanes, voit la diphtongaison de la série des voyelles fermées [e] et [o] comme illustré ci-dessous (cf Cerquiglini, 1995, p. 16).

	pour [e]	pour [o]
Latin	<i>tela, credit, videt</i>	<i>solum, gula, votum</i>
italien	<i>tela, crede, vede</i>	<i>solo, gola, voto</i>
occitan	<i>tela, cre, ve</i>	<i>sol, gola, vot</i>
français	<i>toile, croit, voit</i>	<i>seul, gueule, voeu</i>

Ces changements de voyelles expliquent aussi que par la suite les mots n'ont pas évolué de la même façon en français et dans les autres langues romanes. Alors que les autres langues romanes sont restées très proches du latin, le français a vu la disparition de syllabes entières et bien des racines françaises sont réduites par rapport à la racine du mot latin. Le latin *dormitorium* a donné en italien, en espagnol et en portugais *dormitorio*, alors qu'en français on a *dortoir*, qui ne compte plus que deux syllabes (ibid, p. 17).

Une autre conséquence de cette influence germanique, c'est que le français distingue désormais à l'intérieur des mots des syllabes longues et des syllabes courtes. L'allongement n'est pas la même chose que l'accent tonique ou lexical, qu'on trouve en italien par exemple. En italien, il y a exactement un accent tonique par mot ; en revanche, en français, il n'y a pas nécessairement un allongement par mot. A l'époque médiévale, pour indiquer l'allongement des voyelles accentuées dans la graphie, on a décidé d'utiliser un s. Il s'agissait donc d'un s qui ne se prononçait pas, d'un s qui ne notait pas un son, mais l'allongement de la voyelle qui le précédait. D'où un état de langue, dans lequel la graphie n'est pas étymologisante, mais cherche à transcrire la prosodie de la langue orale. Cela explique pourquoi Charles Beaulieux dans son *Histoire de l'orthographe française* appelle la graphie médiévale la « graphie des chansons de gestes ». C'est une orthographe destinée aux professionnels de l'oral que sont les troubadours et non aux professionnels de l'écrit. Cerquiglini (ibid, p. 39) en parle comme d'une « orthographe non savante ».

Le problème, c'est qu'en utilisant ainsi le s, on introduit une ambiguïté dans le système de l'écriture. Même si on peut dire qu'à cette époque, le s se prononce toujours, il ne se prononce pas toujours de la même façon : tantôt c'est un s qui note un son, tantôt c'est un s qui note un allongement. Pour pallier ce problème, à l'origine de fautes, tant au niveau de l'écriture que de la lecture, certains ont voulu, dans les années 1500, réformer l'orthographe. Les premiers réformateurs, autour de 1540, ont proposé de remplacer le s qui notait l'allongement par un accent, en l'occurrence l'accent circonflexe. De cette manière, on évitait le risque de voir ce s prononcé, un jour, par les ignorants. Au cours des siècles, la prosodie du français a changé et l'allongement a disparu. Du coup, l'accent circonflexe ne servait plus à indiquer une augmentation de la quantité vocalique. En revanche, il a continué à être associé à une variation, mais une variation de timbre et seulement sur certaines voyelles : le e, le a et le o. Les seconds réformateurs, en 1990, ont alors proposé de supprimer l'accent circonflexe sur le i et sur le u, où il ne servait plus à marquer quoi que ce soit. Présentée ainsi, l'histoire de l'accent circonflexe, de son apparition et de sa suppression semble simple. Si elle est en réalité plus compliquée, c'est parce qu'on n'a pas su séparer l'histoire du s graphique dans sa relation à l'accent circonflexe, de l'histoire des autres lettres qui ne se prononcent pas, qu'on appelle des lettres adventices, dans l'écriture du français.

En réalité, à la fin du Moyen-âge, l'orthographe du français est pleine de lettres adventices. On trouve par exemple deux lettres adventices dans le mot *subiect* : un b, qui sert à indiquer que le i est consonantique et se prononce comme l'actuel j (qui ne faisait pas encore partie de l'alphabet) et un c, dont le rôle est de rappeler l'étymon latin de ce mot. On le

voit donc, les lettres adventices remplissent des fonctions très diverses. Elles servent quelquefois à indiquer la prononciation de la lettre qui suit ou qui précède, mais peuvent aussi indiquer des parentés morphologiques en marquant graphiquement l'existence d'une racine commune, comme dans le cas de *fenestre*, dont on veut souligner la parenté avec *défenestrer* (le s de *fenestre* est non prononcé donc adventice). Elles peuvent aussi être utilisées pour distinguer des homonymes : on écrit *poids* avec un d adventice pour éviter la confusion avec *poix* ; de même à côté du mot *mes* on trouve *mets* avec un t adventice ; on utilise aussi un s adventice pour marquer la différence entre l'indicatif et le subjonctif dans les paires verbales *fut* et *fust* et *dit* et *dist*. Deux autres arguments, beaucoup moins rationnels, sont avancés par Estienne pour justifier l'usage de ces lettres qui ne se prononcent pas. Il y a d'une part l'idée qu'avec des lettres adventices et étymologiques, la langue écrite se distingue de la langue orale et gagne en majesté. Le français, langue orale qui se dote d'une langue écrite, peut alors prétendre être une langue de culture, une langue capable de rayonner dans le monde. S'ajoute à cela une dimension sociale : les consonnes étymologiques qui requièrent une connaissance du latin et une longue pratique de l'écrit sont une marque de distinction. Cerquiglini (ibid, p. 68) cite les propos de Bèze, un homme de lettres de l'époque, qui oppose deux groupes sociaux, les gens doctes et les autres : « Il faut qu'il y ait quelque différence entre la manière d'écrire des gens doctes des gens mécaniques » (Bèze, 1584, p. 52). Et il rappelle aussi la position de l'Académie française qui en 1694 tient des propos du même acabit : « il faut suivre l'ancienne manière d'écrire, qui distingue les gens de lettres et qui ont étudié la langue d'avec les ignorants. »

Dès cette époque, il y a eu des débats pour savoir s'il fallait garder ces lettres adventices ou au contraire limiter leur usage. Deux courants s'affrontaient pour ériger les règles d'orthographe. D'un côté un courant phonocentriste, soutenant l'idée d'une reproduction de l'oral dans l'écrit. Ce courant a été défendu en particulier par Ronsard, le poète, dès 1540. De l'autre côté un courant soucieux de maintenir l'histoire, la mémoire dans l'écriture, défendu par l'ensemble des prosateurs (moins soucieux de l'oral que les poètes) au nombre desquels se trouvait Montaigne. Pour comprendre cette seconde position, il faut se replacer dans la situation linguistique de l'époque : les hommes de l'écriture (à savoir, pour l'essentiel, des hommes de droit comme les notaires) sont bilingues, ils parlent à la fois le latin et le français. Les lettres étymologiques ne soulèvent donc aucune difficulté pour eux, et leur pratique quotidienne des deux langues renforce le sentiment d'une parenté entre le français et le latin. Ils ne voient pas pourquoi on devrait effacer les traces de l'histoire de la langue dans l'orthographe, puisque cette orthographe ne leur pose aucun problème : les lettres étymologiques les conduisent à écrire un son qu'ils ne prononcent pas en français, mais qu'ils prononcent en latin. Selon Cerquiglini (ibid, p. 42), « c'est dans ce contexte de bilinguisme culturel qu'il faut comprendre le processus d'étymologisation auquel l'orthographe est soumise dès la fin du XIV^{ème} siècle ». Au XVI^{ème} siècle encore, seul une élite a accès à l'écriture, et cette élite maîtrise parfaitement le latin, donc le recours aux lettres adventices ne constitue pas en tant que tel une difficulté.

Les partisans de l'orthographe réformée et les partisans de l'orthographe ancienne défendent donc des positions inconciliables mais aucun des deux courants n'est libéré de contradiction. Les partisans de l'orthographe réformée reconnaissent les limites d'une orthographe purement phonétique. L'orthographe doit être unifiée, elle doit être la même partout en France, par delà la diversité des dialectes qui existent. Elle ne peut donc figurer qu'une moyenne phonique, ce qui revient à accepter l'idée que l'orthographe est associée à une norme. Quant aux défenseurs de l'orthographe ancienne, ils doivent reconnaître leur incapacité à formuler des règles générales expliquant l'usage des lettres adventices, lettres qui posent une double difficulté, au niveau de l'écriture mais aussi au niveau de la lecture, à tous ceux qui ne maîtrisent pas à la fois le français et le latin.

C'est dans ce contexte que l'introduction de l'accent circonflexe dans la graphie française a été proposée dans les années 1540 : il s'agissait de remplacer par un accent circonflexe les s qui servaient à indiquer l'allongement de la voyelle qui les précédait. Cet accent ne fut adopté par l'académie que 200 ans plus tard, vers 1740. On peut se demander pourquoi et chercher qui étaient venus grossir les troupes des partisans d'une réforme. Il s'agissait en fait essentiellement d'éditeurs (soucieux d'unifier les versions des textes imprimés) mais aussi de pédagogues, et notamment des professeurs enseignant le français aux étrangers. La question de la prononciation des lettres adventices posait un vrai problème à ceux qui voulaient lire ou parler une langue sans être confrontés à son usage, ainsi qu'à ceux qui tentaient de leur enseigner. Devant l'incapacité où se trouvaient les pédagogues à formuler des règles expliquant l'usage des lettres adventices, ils ont milité pour une simplification de l'orthographe française : une orthographe réformée rendrait l'apprentissage du français beaucoup plus facile aux étrangers soucieux de parler et d'écrire cette langue, alors même qu'ils évoluent dans un milieu où ils ne l'entendent pas.

La question de la réforme de l'orthographe est redevenue d'actualité au XX^{ème} siècle. Dans les années 80, une controverse s'est ouverte sur la féminisation des noms de métiers en français et le pouvoir politique a décidé d'intervenir en mettant en place une commission et en s'appuyant sur ses conclusions pour légiférer. La féminisation des noms de métiers est un problème linguistique qui conduit à s'interroger sur l'orthographe : doit-on écrire *professeur* ou *professeure* quand le mot se rapporte à une femme. En 1989, le premier ministre mit en place le Conseil Supérieur de la Langue Française, chargé de proposer une régularisation de l'orthographe sur quelques points comme l'usage du trait d'union, le pluriel des mots composés et bien sûr l'usage du circonflexe. Le but était en réalité principalement de fixer des règles permettant d'encadrer l'écriture des néologismes. Les rectifications furent proposées fin 1990, et il fut bien précisé qu'il ne s'agissait en aucun cas de changer l'orthographe du français mais seulement de la rendre plus permissive. L'Académie Française indique que « l'orthographe actuelle reste d'usage, et [que] les « recommandations » du Conseil supérieur de la langue française ne portent que sur des mots qui pourront être écrits de manière différente sans constituer des incorrections ni être considérés comme des fautes ». Malgré la modestie des recommandations apportées et la faible portée de cette réforme, elle déclencha dans la société de très vifs débats. On retrouve en 1990 les mêmes principes derrière les clivages entre partisans des rectifications et opposants à la réforme. Les partisans défendent l'idée d'une rationalisation de l'orthographe française qui amènerait à diminuer le nombre de règles et à en augmenter la portée, de manière à réduire le nombre d'exceptions à apprendre et à régulariser la graphie des néologismes. Quant aux opposants aux rectifications, ils se focalisèrent sur la question des accents et particulièrement de l'accent circonflexe. Finalement, ce fût au Président de la République de trancher et il résuma son rôle en affirmant qu'« il avait réussi à sauver quelques accents ».

L'histoire de l'accent circonflexe en français est donc emblématique de l'histoire des réformes de l'orthographe en France. Elle est marquée par trois temps : en 1540, les premiers réformateurs proposent de substituer au s qui marque un allongement un accent, le circonflexe, pour éviter de voir se s prononcé à tort ; cette réforme n'est adoptée que 200 ans plus tard ; au XX^{ème} siècle, le phonétisme du français ayant évolué et l'allongement ayant disparu, on propose de supprimer cet accent. Si la réforme a échoué au sens où un nombre important de propositions ont été refusées et où aujourd'hui encore l'emploi de l'orthographe réformée est extrêmement minoritaire, c'est en raison de la pression exercée par ceux que Cerquiglini qualifie de conservateurs modernes et d'anarchistes mondains. Derrière les opposants à la réforme de 1990, écrit-il, « on perçoit également la figure du conservateur moderne. C'est un écrivain, amoureux de sa langue, soucieux de la défendre, empressé qu'on le sache. Il est rétif aux théories, hostiles aux autres confréries du langage surtout quand elles

sont instituées : les linguistes universitaires, qui possèdent un savoir et confèrent des diplômes, l'Académie Française, pour peu qu'elle soit réformatrice et montre par là son autorité [...]. Anarchiste mondain, [...] il marque son indépendance par un goût de ce qui, dans la langue est singulier : [l'accord du participe passé, les pluriels irréguliers]. L'incohérence est belle, [...] l'exception est un joyau rare » (ibid, p. 152-153).

L'histoire de l'accent circonflexe et au delà l'histoire des réformes de l'orthographe montre qu'en ce qui concerne les questions de langue, il faut distinguer savoirs et pratiques. Et cela amène à poser deux questions différentes : qui a le savoir et qui a le pouvoir ? Il arrive que le pouvoir ne soit pas donné à ceux qui ont le savoir. La langue est un objet de sciences, mais c'est aussi un objet historique qui évolue. Et dans la mesure où la langue évolue, les pratiques qui y sont associées ne peuvent qu'évoluer avec elle. De ce fait, ce qui a un moment se présente comme une innovation, peut, quelques siècles plus tard être au contraire une marque de conservatisme. C'est ce qu'illustre bien l'histoire de l'accent circonflexe : ce sont les progressistes qui ont défendu son introduction au XVI^{ème} siècle et aujourd'hui ce sont les conservateurs qui veulent maintenir son usage. Tout est une question de contexte social. Qui était savant, qui ne l'était pas en 1540 ? Cerquiglini rappelle qu'en 1609 les réformateurs récusait le pédantisme graphique au nom du vrai savoir, et tenaient les conservateurs pour de prétentieux demi-savants. Aujourd'hui le contexte social et linguistique a totalement changé : tout le monde apprend à lire et à écrire, mais rares sont ceux qui maîtrisent encore le latin. Qui est savant et qui ne l'est pas ?

On observe dans l'histoire des réformes de l'orthographe une constante : l'existence de deux communautés distinctes qui revendiquent le privilège de la connaissance savante. Il y a d'un côté ceux qui utilisent ou étudient la langue (notamment les poètes au XVI^{ème} siècle et les linguistes aujourd'hui) et de l'autre côté les professionnels de l'écriture. Il semble bien que ce qui a changé, au cours du temps, ce n'est pas tellement l'ensemble de ceux qui étudient la langue, mais bien plutôt l'ensemble de ceux qui ont accès à l'écriture et l'utilise, et parmi eux, la communauté des professionnels de l'écriture. Au Moyen-Age et à la Renaissance, cette communauté était très restreinte et constituée pour l'essentiel de notaires et d'hommes de droit. Aujourd'hui l'enseignement s'étant démocratisé et l'écriture étant désormais accessible au plus grand nombre, cette communauté des professionnels de l'écriture est plus large et elle inclut entre autres les écrivains et les journalistes, souvent très présents dans les médias. Cette évolution des professionnels de l'écriture explique pourquoi et comment l'orthographe s'est réformée. En revanche, la communauté des savants en matière de langue a en un sens moins changé, à la fois qualitativement et quantitativement : les professionnels de la langue sont ceux qui non seulement l'utilisent, mais l'étudie, en synchronie et en diachronie, ils en connaissent l'histoire, mais dissocient la maîtrise d'une langue et la connaissance de son histoire.

L'histoire des réformes de l'orthographe conduit donc à réfléchir à la question des communautés de savoir et à leur légitimité. Quel est ou quels sont les garants de la connaissance savante, qui a la légitimité pour décréter qu'il a le savoir dans tel ou tel domaine ? Mais une autre question se pose, dans la mesure où on a à faire, avec l'orthographe, à une pratique, qui peut être encadrée par des textes de lois. A qui donner le pouvoir de légiférer ? Quand bien même on déciderait de donner ce pouvoir aux savants, il reste à déterminer sur quels principes ces derniers vont fonder la pratique à encadrer. L'histoire de l'orthographe a montré qu'on pouvait légiférer en avançant des principes différents et opter soit pour une pratique efficace, viable économiquement dans une société où l'écriture s'est démocratisée, soit opter pour une pratique savante, réservée à un public instruit. A supposer que les linguistes soient considérés comme les experts pour les questions d'orthographe, ils peuvent ou bien décider d'opter pour une orthographe qui garde la trace de l'histoire de la langue, ou décider d'ignorer la dimension historique et choisir de se placer

dans une démarche entièrement synchronique, où l'on regarde la langue comme un système, qui doit être le plus cohérent possible. Mais ce qui est en tout état de cause illégitime, du point de vue du linguiste qui opte pour une démarche scientifique, qu'elle soit synchronique ou diachronique, c'est d'inventer ou de réécrire l'histoire, comme c'est souvent fait à propos de l'accent circonflexe.

Enfin, un dernier point mérite d'être souligné. Les débats à propos des réformes de l'orthographe ont de tout temps été propices au développement d'idées reçues. Et parmi ces idées, il en est une qui est revenue de manière récurrente : c'est que la langue évolue toujours en s'appauvrissant, qu'avec le temps, elle dégénère. Accepter d'institutionnaliser un usage, ce serait alors toujours accepter d'entériner cette dégénérescence, et au contraire, défendre la langue, ce serait défendre sa richesse et son histoire. On peut parler d'une idée reçue sur la langue parce que rien de scientifique ne permet d'étayer ce propos. On n'a jamais mis en évidence qu'il y avait un appauvrissement de la langue et pour le faire il faudrait d'abord réfléchir à la façon de mesurer la richesse d'une langue : est-elle proportionnelle au nombre de mots du lexique, à la complexité des règles de grammaire utilisées ? Cette idée d'une dégénérescence de la langue va de pair avec une autre idée reçue, selon laquelle il y aurait des langues plus faciles que d'autres, des langues plus logiques que d'autres. En réalité, toute langue naturelle est à la fois simple et complexe. Les langues qui ont une morphologie compliquée, et notamment des conjugaisons et des déclinaisons complexes, sont réputées difficiles. Mais on a observé que la complexité de la morphologie est souvent compensée par une simplicité de la syntaxe. D'aucuns prétendent que certaines langues sont plus faciles à apprendre que d'autres parce qu'elles ont une morphologie pauvre comme l'anglais. Mais le chinois a une morphologie pauvre et qui affirmerait aujourd'hui que c'est une langue facile à apprendre, pour un européen ? Ce qu'on peut mesurer, c'est la distance entre deux langues, mais beaucoup moins facilement la complexité interne de chaque langue. Et le simple fait que tous les enfants du monde mettent sensiblement le même temps à acquérir leur langue maternelle, quelle qu'elle soit, semble indiquer, comme le dit Marina Yaguello (1988, p. 135), qu'un subtil équilibre s'instaure dans toute langue entre le simple et le complexe et que toutes les langues ont sensiblement la même complexité.

Finalement, la question posée par les réformes de l'orthographe est celle de l'institutionnalisation d'une pratique et à travers elle la question de la détermination des communautés savantes garantissant le savoir savant. Si l'orthographe française est pleine d'ambiguïtés et mélange à la fois des principes phonocentriques et des principes étymologiques, si les réformes proposées n'ont jamais vraiment abouti et ne sont pas suivies, c'est sans doute parce qu'on a toujours laissé *in fine* au politique le soin de déterminer quelles réformes faire, alors qu'il ne suffit pas d'écrire des textes de loi pour fixer une norme en matière de langue. On notera que dans un pays comme l'Espagne, où depuis 1713, on a confié à une institution, l'Académie Royale Espagnole, le soin de normaliser la langue espagnole, dans ses usages à la fois écrits et oraux, la situation est fort différente. En matière d'orthographe comme en matière de langue en général, c'est bien plutôt l'usage que la loi qui fixe les pratiques. Imposer une norme qui ne reflète pas l'usage ne peut se solder que par un échec. Le mieux placé pour encadrer une pratique linguistique et infléchir éventuellement l'évolution de la langue, c'est sans doute le linguiste, qui s'est fixé comme tâche d'observer les usages, de décrire leur évolution et de mettre à jour les lois permettant d'expliquer ces évolutions.

3) LA QUESTION DE L'ACCES AUX DONNEES LINGUISTIQUES : INTROSPECTION, TRAVAIL SUR CORPUS ET EXPERIMENTATION

Dans son livre *Introduction à une science du langage*, Milner défend l'idée que la linguistique est une science expérimentale au sens où le linguiste se donne pour tâche de formuler des hypothèses sur la langue, qu'il va vérifier ou falsifier en les confrontant aux faits. Un des apports de la grammaire générative est d'avoir montré qu'on pouvait tester une hypothèse linguistique en confrontant les prédictions qu'elle permet de faire à la réalité des faits grammaticaux. Ces prédictions prennent la forme de paires minimales, qui opposent une forme grammaticale à une forme agrammaticale. Ainsi par exemple on peut dire qu'en français le sujet s'accorde avec le verbe, ce que corrobore la paire minimale *Les hommes ont faim* / * *Les hommes a faim*. Le linguiste doit donc s'intéresser non seulement aux énoncés bien formés dans une langue donnée, mais aussi être capable de prédire quels énoncés sont mal formés, et pour noter cette agrammaticalité il utilise le symbole *. La question qui se pose est alors de savoir sur quelle base on décide qu'un énoncé est grammatical ou agrammatical, qu'il fait partie du système de la langue ou en est exclu. Chomsky a défendu l'idée qu'on pouvait s'appuyer sur la compétence des locuteurs natifs et que c'est par introspection qu'un locuteur natif pouvait déterminer si tel énoncé était grammatical ou non. Il a donc donné une place centrale aux jugements introspectifs en linguistique. De deux choses l'une : ou le linguiste travaille sur sa langue maternelle et dans ce cas il peut utiliser son propre jugement introspectif pour vérifier ses hypothèses, ou il travaille sur une langue qui lui est étrangère et dans ce cas il lui suffit d'interroger un locuteur natif, qui lui sert d'informateur.

Le recours à l'introspection comme principe méthodologique a souvent été mis en cause. L'expérience a montré qu'on ne pouvait pas toujours se fier aux jugements des locuteurs natifs. En effet, il n'est pas rare qu'un locuteur natif soutienne, quand on l'interroge, que telle ou telle expression est agrammaticale, alors qu'une recherche sur corpus permet d'en trouver plusieurs occurrences. C'est que si les locuteurs natifs se prononcent facilement sur la grammaticalité d'une phrase, en revanche, ils se montrent souvent incapables de répondre quand on leur demande dans quel contexte un énoncé particulier pourrait apparaître. Il est fréquent qu'un locuteur juge agrammatical ce qui, dans les faits, s'avère être un énoncé attesté. Mais le locuteur en question reconnaît que cet énoncé est grammatical quand il est replacé dans son contexte. Il y a donc une instabilité du jugement de grammaticalité fondé sur l'introspection des locuteurs natifs, qui vient de ce qu'il est quelquefois difficile d'imaginer le contexte qui légitimerait l'occurrence d'un énoncé particulier.

Une seconde critique contre les jugements de grammaticalité et l'introspection a été formulée par Schutze (1996) qui soutient, exemples à l'appui, que le jugement de grammaticalité n'est pas catégoriel, qu'il est des cas limites pour lesquels les locuteurs hésitent. Il en va ainsi de l'accord au singulier ou au pluriel du verbe quand le sujet est une disjonction. Schutze donne l'exemple de la question *??Are / ??Is John or his parents here ?* où les points d'interrogation devant les deux formes verbales indiquent la difficulté des locuteurs à se prononcer. On trouve des données comparables en français.

Un des problèmes mis en avant dans les débats récents sur l'usage de l'introspection en linguistique, c'est qu'à la différence de ce qui s'est passé aux premiers temps de la grammaire générative, aujourd'hui les théories s'intéressent à des questions très complexes et se développent sur la base d'observations qui sont de moins en moins fiables. Ce problème avait été anticipé dans les années 1970 par Levelt qui écrivait : « It has slowly but surely become clear that it is not possible, on the basis of incontrovertible, directly evident data, to construct a theory so extensive that all less obvious cases can be decided upon by the grammar itself. It is becoming more and more apparent that decisions on very important areas of theory are dependent on very unreliable observations » (Levelt 1974, vol. 2, p. 6).

Enfin le dernier point est que le jugement introspectif n'est pas falsifiable. En effet, que peut-on opposer à un locuteur qui soutient que tel ou tel énoncé lui semble grammatical ? Uniquement le jugement d'un autre locuteur, ce qui ne prouve rien, sinon peut-être que ces

deux locuteurs ne partagent pas la même grammaire, ne parlent pas, en un sens, la même langue. D'une certaine façon, l'introspection donne accès à une information proprement subjective, qu'il est difficile d'objectiver.

On voit donc que le recours à l'introspection en linguistique, pour établir les données sur la base desquelles construire une théorie, fait débat. Qui plus est, il apparaît que pour certaines questions linguistiques, l'introspection n'est d'aucun secours. C'est le cas notamment quand on travaille en diachronie ou sur une langue morte, mais aussi quand on s'intéresse comme Bresnan (2007) à des questions de préférences. Bresnan a travaillé sur l'alternance dative en anglais : elle part de l'observation qu'on peut en anglais construire un verbe comme *to give* avec deux objets directs (*to give John a book*) ou avec un objet direct et un objet prépositionnel (*to give a book to John*). Son but est de montrer que même si ces deux constructions sont possibles, elles ne correspondent pas à des variantes libres, au sens où l'une et l'autre seraient utilisées absolument indifféremment, mais qu'au contraire il existe certains critères (comme la nature nominale ou pronominale de ces compléments, leur longueur, le fait qu'ils soient définis ou indéfinis...) qui font que dans un contexte particulier, les locuteurs manifestent une préférence nette pour l'emploi d'une de ces deux formes par rapport à l'autre. L'introspection ne permet pas d'avoir accès à ces informations sur les préférences. En revanche, d'autres méthodes peuvent permettre de les mettre à jour.

Pour tester la validité d'une hypothèse linguistique, on peut donc utiliser d'autres méthodes que celle qui repose sur le jugement introspectif. On peut trouver une description assez complète de ces méthodes, ainsi que des avantages et des inconvénients inhérents à chacune d'elles dans Meyer et Nelson (2006). On peut par exemple travailler sur des enregistrements audio et video de conversations réelles, ou exploiter des corpus de données orales ou écrites et en faire l'analyse, quantitative et qualitative. On peut aussi, pour vérifier une hypothèse donnée, utiliser les méthodes de la psychologie expérimentale et demander à des sujets d'accomplir une tâche langagière particulière, qui peut être soit une tâche de production (on demande alors aux sujets ce qu'ils diraient ou écriraient dans tel contexte) ou une tâche d'interprétation (on demande aux sujets comment ils interprètent un énoncé donné placé dans un contexte particulier). Dès qu'on travaille sur des corpus, que ce soient des corpus généraux ou des corpus construits pour tester une hypothèse particulière, on analyse ces corpus au moyen de méthodes statistiques. Les statistiques peuvent servir soit à décrire des faits de langue (qui ne sont pas nécessairement catégoriels, mais peuvent correspondre à des préférences) soit pour décider de la significativité de certaines différences et généraliser une observation au delà d'un échantillon donné. On fait alors des statistiques inférentielles. On peut aussi utiliser les statistiques pour découvrir des propriétés cachées dans les données. Dans ce cas, on fait de la modélisation statistique et on utilise en particulier les modèles de régression.

Il n'y a pas de raison d'opposer ces méthodes entre elles. Au contraire, elles peuvent se compléter, voire même se conjuguer. C'est le point de vue que défendent Gibson et al. (2010, 2012) dans le débat récent sur la place de l'introspection en linguistique, qui les opposent notamment à Sprouse et Almeida (2012). Dans un premier article, Gibson et Fedorenko (2010) mettent en cause l'introspection comme principe méthodologique pour fonder les recherches en syntaxe et en sémantique. Ils expliquent qu'il n'y a aucune raison aujourd'hui qui justifient que les linguistes n'aient pas recours à d'autres méthodes que la simple introspection pour collecter les données sur lesquelles ils fondent leurs théories. S'il fut un temps où les seules expériences qu'il était raisonnable de conduire étaient des expériences de pensée, tel n'est plus le cas. En effet, on peut très facilement en utilisant internet et la plateforme *Mechanical Turk* par exemple conduire une expérience en ligne et recueillir en très peu de temps les jugements d'une grande quantité de locuteurs natifs, ce qui permet d'objectiver les données et de mener des études quantitatives sur des questions linguistiques.

Après avoir rappelé les limites de l'introspection telle qu'elle est pratiquée par la plupart des linguistes, Gibson et Fedorenko soutiennent que contrairement à ce qui est affirmé par beaucoup de linguistes et notamment par Phillips (2009), on trouve dans la littérature en syntaxe et en sémantique des exemples de jugements introspectifs incorrects qui ont servi de fondement au développement d'une théorie douteuse ou à l'expression d'une généralisation invalide. Ils développent dans le détail trois exemples (le premier concerne les relatives doublement enchâssées et leur complexité selon qu'elles modifient un sujet ou un objet, le second porte sur l'asymétrie sujet / objet dans les questions multiples du genre *Who bought what?* / **What did who buy?* et le troisième concerne l'extraction de mots interrogatifs dans les questions multiples) et ils expliquent que pour éviter ce type d'écueil, il serait important que de façon systématique les recherches en syntaxe et en sémantique prennent appui sur quelques données quantitatives. Selon eux, si les linguistes prenaient l'habitude de vérifier leurs intuitions en les confrontant soit à une recherche sur corpus soit aux résultats d'expériences menées auprès d'un nombre significatif de locuteurs, leurs travaux recevraient plus d'écho dans des communautés où l'expérimentation est systématique et toujours rigoureusement contrôlée. Sprouse et Almeida ont réagi à cette attaque contre les partisans de l'introspection et ont testé, en utilisant les méthodes quantitatives préconisées par Gibson et Fedorenko, la fiabilité des exemples utilisés dans un des manuels de syntaxe les plus utilisés (*Core Syntax*, de David Adger, publié en 2003) ainsi que l'ensemble des exemples cités dans les articles de la revue *Linguistic Inquiry*, parus entre 2001 et 2010. Ils ont constaté que dans 95% des cas, les jugements obtenus en utilisant une méthode expérimentale étaient identiques aux jugements obtenus par introspection. Ce qui les amène à conclure que les données utilisées par les syntacticiens ayant recours à l'introspection pour fonder leurs théories sont extrêmement fiables et qu'il n'y a pas de réelles raisons pour que les linguistes changent de méthodes. Gibson, Piantadosi et Fedorenko (2012) ont alors répondu que l'étude faite par Sprouse et Almeida apportait en réalité de l'eau à leur moulin. D'une part, elle montrait bien qu'il était facile aujourd'hui de confronter les données de l'introspection à des résultats expérimentaux, puisque Sprouse et Almeida avait pu en très peu de temps tester tous les exemples de l'ensemble des articles parus ces dix dernières années dans une grande revue. D'autre part, selon eux, le taux de 95% obtenu par les auteurs n'est pas un résultat très satisfaisant. En effet, cela prouve qu'il y a 5% des exemples utilisés par les syntacticiens qui ne sont pas fiables et il faudrait vérifier quel rôle jouent précisément ces exemples dans l'élaboration des généralisations défendues dans ces articles. Il est tout à fait possible, selon Gibson et al., qu'il s'agisse justement des exemples qui jouent un rôle crucial pour élaborer les hypothèses théoriques défendues dans les articles.

Du débat qui oppose les linguistes sur la place de l'introspection en linguistique, il ressort qu'on ne peut ni rejeter l'introspection en bloc, ni non plus en faire le seul mode d'investigation de la langue, comme le préconisait Chomsky (1965, p. 20) quand il soutenait qu'il était inutile d'essayer d'affiner les données sur lesquelles travailler en utilisant d'autres méthodes plus objectives pour les tester. L'introspection ne peut plus être l'unique méthode utilisée par le linguiste, c'est une des méthodes utilisées par le linguiste pour collecter des données. Il faut toutefois noter que seule l'introspection ou des expériences testant directement l'acceptabilité des exemples permet de faire des prédictions négatives et d'exploiter les contrastes entre données grammaticales et données agrammaticales. Le fait qu'un type d'exemples n'apparaisse pas dans un corpus ne peut pas être utilisé comme signifiant que ce type de données est agrammatical. En un sens donc, le jugement fondé sur l'introspection est quelquefois incontournable. Mais parce que ce jugement peut être trompeur, il est utile d'avoir recours à d'autres méthodes que la simple introspection individuelle pour collecter des données. La solution ne consiste donc pas à rejeter ce mode de production des données, mais au contraire à mieux le contrôler et à en connaître les limites,

car sans introspection (et par là nous entendons non seulement l'introspection du linguiste lui-même, mais les jugements d'acceptabilité des témoins qu'il interroge), il n'est pas d'études sérieuses possibles de la syntaxe et de la sémantique. Le second point intéressant, c'est que l'introspection, qui à l'origine était une méthode de recherche pour le linguiste, est devenue maintenant un objet de recherche.

Parce que chaque individu utilise le langage au quotidien, il peut prétendre avoir une certaine connaissance de sa langue. S'agit-il d'une connaissance vulgaire, qui se distinguerait de la connaissance savante des linguistes ? Il y a en effet une différence de nature entre la connaissance que chacun a de sa langue et la connaissance des linguistes, qui s'intéressent non seulement aux productions individuelles, mais aussi au fonctionnement général d'une langue donnée, à la place que cette langue occupe dans l'ensemble des langues du monde, et plus généralement à la faculté de langage et aux universaux linguistiques, dont chaque langue est une réalisation particulière. Cependant, si on considère la relation de chaque individu à sa propre langue, il faut encore distinguer deux niveaux d'appréhension différents, qui correspondent sans doute à deux types de connaissances vulgaires différentes : il y a d'une part la langue telle que je la parle, et d'autre part la langue telle que je devrais la parler, celle qu'on m'a apprise, en particulier à l'école, avec un premier décalage entre usage effectif et usage normé. Si on voit dans la compétence linguistique de chaque individu une forme de connaissance vulgaire et dans la connaissance des linguistes une forme de connaissance savante, le débat sur l'introspection qui agite la discipline aujourd'hui semble montrer qu'en linguistique comme en économie « l'opposition entre les deux formes de savoir se réduit par une empiricisation croissante » de la discipline, pour reprendre les termes de Walliser dans ce même volume.

4. Connaissance vulgaire et idées reçues

On sait peu de choses des connaissances vulgaires en général. Mais si la connaissance vulgaire, c'est la connaissance de sens commun, celle qui est partagée par le plus grand nombre, peut-être fait-elle une place aux idées reçues. Or il existe une littérature qui se donne pour tâche de recenser et de combattre les idées fausses et les idées reçues dans tel ou tel domaine. Mais que sont ces idées reçues, si ce n'est des idées un peu générales, souvent approximatives, dont certaines sont inculquées à l'école et d'autres véhiculées par la société ? Cependant, il faut distinguer plusieurs types d'idées reçues. Il y a les idées reçues qui correspondent à des idées fausses. Ce ne sont tout simplement pas des connaissances, elles ne font donc pas parties de la connaissance vulgaire. Il y a aussi les idées reçues qui ne sont pas à proprement parler des idées fausses, mais plutôt des simplifications de la connaissance savante. Beaucoup d'entre elles sont véhiculées par l'école, car on ne peut enseigner que progressivement et la pédagogie consiste souvent à simplifier les choses dans un premier temps pour n'en montrer toute la complexité que dans un second temps. Enfin, il y a les lieux communs, issus souvent de l'expérience subjective de tout un chacun mais partagés, connus de tous, dont on verra qu'ils peuvent servir de points de départ à des recherches savantes, ces dernières se fixant comme objectif soit de les vérifier soit de les falsifier.

Essayons d'illustrer ces différents types de savoirs ou connaissances. Les idées reçues sur la langue française sont légion. Sont régulièrement publiés des ouvrages de linguistes qui en recensent quelques-unes pour les démontrer. Récemment encore est sorti un ouvrage de ce type (Rittaud-Hutinet, Chantal, *Parlez-vous français ? Idées reçues sur la langue française*, 2011), ouvrage qui s'inscrit dans la même veine que le travail plus ancien de Marina Yaguello, *Catalogue des idées reçues sur la langue*, paru en 1988. Au nombre de ces lieux communs, il y a par exemple les affirmations suivantes :

- Le français se dégrade, s'appauvrit.

- La structure sujet-verbe-complément représente l'ordre normal de la phrase en français.
- Certaines langues sont plus logiques que d'autres.
- Une langue pure est parlée sans accent.
- Un dialecte est une sous-langue, sans grammaire et sans écriture.
- Un enfant qui apprend deux langues à la fois ne se sentira jamais à l'aise dans aucune.
- Les langues "primitives" sont simples.

On ne va pas ici réfuter ces idées reçues, mais plutôt essayer d'en faire une typologie, en montrant que toutes n'ont pas la même source.

Comme on l'a dit, certaines de ces idées reçues sont le résultat de la simplification d'une connaissance savante. C'est le cas par exemple de l'affirmation selon laquelle la structure sujet-verbe-complément représente l'ordre normal de la phrase en français. Cette affirmation n'est pas à proprement parler fautive, mais bien vague, la question étant de déterminer le sens de « normal » dans l'expression « l'ordre normal ». Ce qu'ont observé les grammairiens et qu'ont établi les linguistes, c'est qu'on peut définir ce qu'on appelle la phrase canonique du français, qui correspond à une phrase déclarative minimale, du type *Jean aime Marie* et dans laquelle l'ordre des mots est bien sujet-verbe-complément. Et on peut ensuite établir un ensemble de règles qui permettent de montrer comment dériver, à partir de cette phrase canonique, par transformation par exemple, toutes les autres phrases du français. Mais la phrase canonique est un objet de pensée, qui ne correspond à aucune norme, pas plus qu'elle ne correspond à la phrase-type ou la phrase la plus fréquente du français. On ne peut dire ni que toutes les phrases du français suivent l'ordre sujet-verbe-complément, ni même que la majorité d'entre elles ont cette structure.

Il est d'autres idées reçues qui véhiculent une idéologie. Marina Yaguello soutient que c'est au linguiste de démontrer ces idées reçues et que ce travail est important à deux titres : pour rétablir la vérité, et parce que le plus souvent ces idées reçues sont loin d'être innocentes, elles sont empreintes d'idéologie, et peuvent « nuire à la compréhension de l'autre, donner des arguments à toutes les formes de racismes et contribuer à l'obscurantisme » (Yaguello, 1988, p. 14). Les stéréotypes et les caricatures contribuent à la diffusion de ce type d'idées reçues. Selon elle, c'est depuis Tintin au Congo que le stéréotype du Noir qui ne prononce pas les r s'est imposé dans les bandes dessinées, la publicité et les films pour enfants... On retrouve ce même trait (l'absence de r-) dans les caricatures de l'accent africain ou antillais. Et de là serait né le préjugé selon lequel les noirs seraient génétiquement inaptes à prononcer le r. (Yaguello, 1988, p. 43).

On peut donner un autre exemple d'idée reçue qui mêle déformation d'une connaissance savante et idéologie. C'est celle qui concerne l'opposition entre langue et dialecte. Un gouffre sépare sur ce point les conceptions du commun des mortels de l'approche du linguiste. Le linguiste donne deux définitions de la langue, l'une linguistique et l'autre sociolinguistique. « Du point de vue linguistique, la langue est l'ensemble de tous les dialectes, répartis dans l'espace social ou régional, qui assurent une intercompréhension suffisante entre leurs locuteurs respectifs. La langue française est donc la somme des dialectes dont les locuteurs se reconnaissent eux mêmes comme francophones natifs, par delà les classes sociales et la variation régionale » (Yaguello, 1988, p. 34-35). Cette définition se fonde sur la notion de continuum dialectal. La différenciation est graduelle et c'est seulement aux extrémités du continuum, lorsque l'intercompréhension n'est plus assurée qu'on peut dire qu'on est passé d'une langue à une autre.

À la différence du linguiste, le non-linguiste introduit en général une hiérarchisation entre langues et dialectes, parle de « grandes langues » et de « petites langues », assimile souvent langues à langues nationales et dialecte à langues régionales. Il assimile souvent différence de statuts (langue plus ou moins dominante) avec une différence de valeur interne.

Or le statut d'une langue résulte de circonstances qui n'ont rien à voir avec sa nature (Ibid, p. 37-38).

Enfin, certaines idées reçues viennent directement de l'expérience, sans s'appuyer sur aucune connaissance savante. Elles peuvent alors servir de point de départ à des recherches nouvelles, dont le but est de les mettre à l'épreuve, pour les vérifier ou les falsifier. Il ne faut donc pas penser que le mouvement va toujours de la connaissance savante vers la connaissance vulgaire. Les observations sur le bilinguisme ou le plurilinguisme en sont un exemple. Elles ont donné lieu au développement de recherches sur l'acquisition des langues, tant de la langue première que de la langue seconde.

On entend souvent dire que plus on parle de langue, plus il est facile d'en apprendre de nouvelles, et parallèlement, on entend aussi souvent dire que les bilingues ne maîtrisent parfaitement aucune des deux langues. Des recherches ont été menées pour tester ces deux hypothèses et ont conduit à préciser les questions, en mettant en avant des dimensions pertinentes et non prises en compte comme le fait que les langues maîtrisées ou à apprendre soient ou non apparentées, la richesse comparée des systèmes phonologiques de la langue première et de la langue seconde... On a pu montrer par exemple, que si l'apprentissage d'une langue n'est pas qu'une question d'oreille, la richesse phonologique de la langue première est un facteur important dans l'apprentissage pour percevoir les sons distinctifs de la langue seconde qui sont absents de la langue maternelle. Cette dimension peut expliquer pourquoi les Russes par exemple ont des facilités à apprendre les langues étrangères et qu'inversement, les Français, dont la bande de fréquence est plus élevée, parviennent moins bien à apprendre l'anglais (cf. Yaguello, p. 32). Tout locuteur approche une langue étrangère à partir du filtre de sa propre langue maternelle.

En conclusion, il semble que les idées reçues sont d'autant plus nombreuses que la langue est le bien commun de tous, que chaque locuteur est un sujet parlant qui se fait une certaine idée de la langue, idée qu'il traduit souvent par un jugement de valeur. Les linguistes tentent régulièrement de rectifier ces discours, mais peinent à convaincre et force est de constater que sur les questions de langue, les impressions prennent souvent le pas sur les faits.

Conclusion

Les exemples étudiés ont montré que même s'il existe une forme de porosité entre connaissance savante et connaissance vulgaire, pour ancrer ce qui différencie ces deux types de connaissance, on peut mettre en avant plusieurs propriétés : la connaissance vulgaire est produite pour le vulgaire, à destination du vulgaire par opposition à la connaissance savante produite par et pour les savants (éventuellement pour les savants d'autres disciplines, comme dans le cas des connaissances sur les pathologies du langage par exemple) ; la connaissance vulgaire est tournée vers une pratique, à la différence de la connaissance savante qui comporte toujours un aspect théorique ; la connaissance savante est réflexive (au sens où elle s'interroge sur son objet), elle est argumentée, et systématiquement validée par une communauté de savants qui s'en portent garants ; parce qu'elle est plus abstraite que la connaissance vulgaire, la connaissance savante construit et manipule des concepts, elle s'énonce dans une langue savante, un « métalangage », alors que la connaissance vulgaire reste très concrète et ne fait pas appel à autre chose que le langage ordinaire quotidien pour s'énoncer.

Un autre point mérite d'être souligné, qui concerne spécifiquement l'opposition entre connaissance savante et connaissance vulgaire dans le domaine du langage. Les linguistes ont toujours tenus à se libérer de l'idée de norme linguistique : ils étudient la langue telle qu'elle est, décrivent son évolution, mais ne s'intéressent pas, à la différence du vulgaire, à la langue telle qu'elle devrait être. Le savoir du vulgaire sur la langue, c'est toujours un savoir qui porte sur la belle langue, le bien parler, un savoir qui s'appuie sur des règles qui dictent comment on doit parler. De ce point de vue, le vulgaire et le grammairien entretiennent le même type de

rapport à la langue, un rapport normatif. Or si de tout temps la norme linguistique a intéressé le grammairien, par principe le linguiste l'a toujours rejetée. En tant que discipline scientifique, la linguistique s'interdit de porter des jugements de valeur. Dans son rapport à la langue et à la norme, le vulgaire est plus proche du grammairien que du linguiste. Cela vient en partie de ce que l'école se charge d'encadrer l'enseignement de la langue (notamment de la langue écrite) et de ce que tout écolier reçoit des leçons de grammaire qui façonnent directement l'image qu'il se fait de sa langue. On devra donc distinguer au moins trois types de connaissance : la connaissance vulgaire orientée vers la pratique, la connaissance savante des grammairiens qui cherchent à encadrer cette pratique, et la connaissance savante des linguistes qui est à la fois descriptive et théorique, mais jamais normative.

Enfin, il nous semble que l'opposition entre savoir savant et savoir vulgaire ne s'articule pas de la même façon dans les sciences humaines et dans d'autres domaines. Alors qu'en physique ou en mathématique, les savants appuient leur connaissance sur des preuves, rien de tel n'existe en sciences humaines. Au travers des exemples que nous avons étudiés, il est apparu qu'en réalité, les arguments avancés par les savants articulaient toujours une connaissance de l'objet étudié et une connaissance de son histoire. Il semble donc qu'un lien étroit doive être établi dans les sciences humaines entre connaissance savante et histoire des sciences.

Références

- Auroux Sylvain, *La question de l'origine des langues*, Paris : PUF, 2007.
- Beaulieux Charles, *Histoire de l'orthographe française*, Paris : Champion, 1927.
- Bèze Théodore (de), *De Francicae linguae recta pronuntiatione*, Genève : Vignon, 1584.
- Bresnan Joan, Is syntactic knowlege probabilistic ? Experiments with the English dative alternation, in Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in Search of Its Evidential Base*, Berlin: Mouton de Gruyter, 77-96, 2007.
- Catach Nina, *L'orthographe française à l'époque de la renaissance : auteurs, imprimeurs, ateliers d'imprimerie*, Genève : Droz, 1968.
- Catach Nina, *L'orthographe en débat. Dossiers pour un changement*, Paris : Nathan, 1991.
- Cavalli-Sfoza Luigi Luca, *Gènes, peuples et langues*, Paris : Odile Jacob, 1996.
- Cerquiglini Bernard, *Charte européenne des langues régionales ou minoritaires. Rapport explicatif*. Les éditions du Conseil de l'Europe, 1992.
- Cerquiglini Bernard, *L'accent du souvenir*, Paris : Les éditions de Minuit, 1995.
- Chomsky Noam, *Aspects of Theory of Syntax*, Cambridge : MIT Press, 1965.
- Court de Gébelin Antoine, *Monde primitif analysé et comparé avec le monde moderne*, Paris : l'auteur, 9 vol in-4°, 1773-1782.
- Estienne Robert, *Dictionnaire francoislatin, autrement dict, les mots francois avec les manieres duser diceulx, tournez en latin*, Paris : Robert Estienne, 1539.
- Gibson Edward & Federenko Evelina, The need for quantitative methods in syntax/semantics research, *Language and Cognitive Processes*, DOI :10.180/01690965.2010.515080, 2010.
- Gibson Edward, Piantadosi Steven T. & Federenko Evelina, Quantitative methods in syntax/semantics research : a response to Sprouse and Almeida (2012), *Language and Cognitive Processes*, DOI :10.180/01690965.2012.704385, 2012.
- Greenberg Joseph H., *Language in the Americas*, Stanford : Stanford University Press, 1987.
- Hombert Jean-Marie, (ed.), *Aux Origines des langues et du langage*, Fayard, 2005.
- Levelt Pim, *Formal grammars in linguistics and psycholinguistics*. The Hague: Mouton, 1974.
- Mc Grew William, *Chimpanzee material Culture : implications for human evolution*, Cambridge : Cambridge Universiry Press, 1992.

- Meyer Charles F. & Nelson Gerald, Data collection, in Bas Arts & April McMahon (eds.), *The Handbook of English Linguistics*, chapitre 5, Malden MA : Blackwell, 93-113, 2006.
- Milner Jean-Claude, *Introduction à une science du langage*, Paris : Seuil, 1989.
- Picq Pascal, Sagart Laurent, Dehaene Ghislaine, Lestienne Cécile, *La plus belle histoire du langage*, Paris : Seuil, 2008.
- Phillips Collin, Should we impeach armchair linguists? in S. Iwasaki, H. Hoji, P. Clancy, & S.-O. Sohn (eds.), *Japanese-Korean Linguistics* Vol. 17, Stanford, CA: CSLI Publications, 2009.
- Pruvost Jean, *Dictionnaires et nouvelles technologies*, Paris, PUF, 2000.
- Pruvost Jean, *Les dictionnaires de la langue française : une histoire et une dynamique*, Musée virtuel des dictionnaires, <http://www3.u-cergy.fr/dictionnaires/index.html>, 2002.
- Rey Alain, *Le Lexique : images et modèles*, Paris : Armand Colin, 1977.
- Rey Alain, *Encyclopédies et dictionnaires*, Paris, Que sais-je ?, 1982.
- Rey-Debove Josette, *Étude linguistique et sémiotique des dictionnaires français contemporains*, The Hague, Paris, 1971.
- Richelet Pierre, *Dictionnaire français*, Genève : J.H. Widerhold, 1680.
- Ringe, Donald A., On calculating the factor of chance in language comparison, *Transactions of the American Society* 82, 1-110, 1992.
- Rittaud-Hutinet Chantal, *Parlez-vous français ? Idées reçues sur la langue française*, Le Cavalier Bleu Éditions, 2011.
- Ruhlen Merritt, *The Origin of Language. Tracing the Evolution of the Mother Tongue*, New-york : John Wiley & sons Inc., 1994.
- Schutze Carson T., *The empirical base of linguistics : grammaticality judgments and linguistic methodology*, Chicago : University of Chicago Press, 1996.
- Sprouse Jon & Almeida Diogo. The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes*, DOI: 10.1080/01690965.2012.703782, 2012.
- Swadesh Morris, *The origin and diversification of languages*, Routledge & Kegan Paul, Londres, 1972. Traduction française : *Le langage et la vie humaine*, Paris : Payot, 1987.
- Yaguello Marina, *Catalogue des idées reçues sur la langue*, Paris : Points Seuil, 1988.