

Claude Hagège

## L'informatique et la langue

*Secousse : Dans quelles mesures les outils informatiques facilitent-ils les études comparatives entre langues et familles de langues ? Ont-ils permis d'améliorer notre compréhension de leur évolution historique ?*

CH : Sans aucun doute. Quand nous comparons des langues entre elles, génétiquement (c'est-à-dire par famille de langues) ou typologiquement (c'est-à-dire par types de langue), nous devons brasser un nombre de données considérable. Grâce aux outils informatiques, nous avons le moyen de réunir ces données sans avoir à les chercher dans les ouvrages écrits et à les compiler au prix de très longues investigations. Cela nous rend un service immense. Les études linguistiques ont été fortement servies par l'informatique.

Prenons l'exemple des langues africaines. Il en existe un nombre très important. Jusqu'ici, il était long et malaisé de les répartir en familles. Grâce à l'ordinateur, nous avons un accès immédiat à tout le corpus. Cela suppose évidemment un travail préalable. Les services que nous rend l'ordinateur sont limités par le fait qu'il faut au préalable le conditionner, entrer les informations nécessaires – il ne peut pas les inventer. Mais une fois qu'elles sont enregistrées, elles sont facilement accessibles. Nous pouvons ainsi retrouver, par un simple clic, les données sur les structures grammaticales ou sur les relations lexicales entre deux langues (il suffit d'entrer le mot pertinent), ce qui demandait autrefois beaucoup plus de temps. Sur ce plan, dans le travail comparatiste, en linguistique comme dans la plupart des disciplines où l'on compare (comparer, c'est quasiment l'une des définitions de la recherche scientifique), l'ordinateur est un auxiliaire très important.

Autre exemple : les langues tibéto-birmanes : le chinois, le birman, le lolo, le haka, le tibétain, diverses autres langues de l'Asie et de l'Himalaya. Jusqu'à présent, les tibéto-birmanologues étaient très embarrassés. Ces langues sont un ensemble confus, très complexe. À partir du moment où ils ont les moyens d'appeler tout de suite une langue, en fonction d'une autre qu'ils connaissent déjà, les tibéto-birmanologues y voient beaucoup plus clair. On peut ainsi présenter l'ensemble des langues, avec les différentes subdivisions – toutes choses que l'on a faites soi-même. L'ordinateur, en somme (je propose une définition qui a peut-être son intérêt), est un instrument qui restitue en un temps extrêmement court les connaissances qu'on a mis un temps considérable à lui donner : mais en les classant et en les ordonnant.

*Secousse :... Est-ce que les analyses informatiques ont permis d'acquérir des certitudes sur l'existence de ce que l'on appelle l'indo-européen ?*

CH : Non, elles n'ont pas permis de conclure. Il n'y a pas d'autres façons de comparer les langues apparentées sous le nom de la famille indo-européenne que d'avoir

d'énormes corpus de langues. Or, s'agissant de langues souvent très anciennes, parfois disparues depuis des millénaires, comme le hittite, le tokharien, c'est un travail très difficile. Il faut souligner que, lorsqu'on le nourrit par un grand nombre de données, l'ordinateur peut les brasser et accélérer les analyses ; mais le processus de création, d'invention, de découverte, est jusqu'ici humain et uniquement humain. Malgré son caractère d'auxiliaire capital, malgré la possibilité qu'il offre de vérifier rapidement beaucoup d'hypothèses, l'ordinateur n'a pas jusqu'ici pour effet la découverte. Cela pourrait changer, en particulier grâce aux robots – je pense à *Pepper*, un robot dont on parle beaucoup, dont on pense qu'un jour il pourrait devenir amoureux... L'ordinateur a contribué à mieux faire connaître les affects humains. Mais ira-t-il plus loin ? On n'en sait rien.

*Secousse : L'ordinateur, mais également les autres outils informatiques (internet, tablettes, réseaux sociaux, blogs, etc.), sont-ils la cause d'une évolution significative de la langue française, orale et écrite ? Ont-ils modifié notre rapport avec la langue ?*

CH : Jusqu'ici non. Je ne crois pas que les outils informatiques aient eu un effet particulier sur l'accélération ou le ralentissement de l'évolution de la langue française, ni d'ailleurs des autres langues. La tendance que l'on observe à l'oral, à aller vite, à ne pas construire de phrases, à faire des blocs de mots, à raccourcir ces derniers, etc., cette tendance est permanente. Elle est même historique. Que ce soit par le moyen de l'ordinateur, aujourd'hui, ou bien, autrefois, par certains modes expéditifs de communiquer, rien n'a vraiment changé. On ne peut pas dire que l'évolution de la langue soit plus forte aujourd'hui, du fait de l'outillage informatique, qu'elle ne l'était autrefois.

Quant au fait de savoir si l'ordinateur aurait changé notre rapport à la langue, ma première réaction serait de dire non, mais il faudrait examiner cela de plus près. Je ne vois pas comment les outils informatiques auraient pu la modifier. L'ordinateur n'a pas eu d'effet sur une chose capitale, qui est la différence entre l'écrit et l'oral. Notre rapport à la langue, surtout pour une langue très normée comme la langue française, est essentiellement un rapport entre deux normes différentes et quelquefois divergentes. Quand vous écrivez un texte pour être embauché, vous ne pouvez pas écrire comme vous parlez. Et quand vous parlez à un ami de trente ans, vous ne lui parlez comme vous écrivez. En d'autres termes, ces différences fondamentales de l'écrit et du parler n'ont pas été modifiées par l'ordinateur, car elles lui préexistaient, et si elles semblent s'accélérer aujourd'hui, c'est du fait de l'extension des communications.

*Secousse : L'un des champs actuels de recherche est la « linguistique informatique ». Pourriez-vous nous décrire simplement ses objectifs, ses méthodes, ses résultats ?*

CH : Il y a un type de linguistique, que certains peuvent en effet appeler *informatique* et qu'on appellera plutôt *formelle*, qui consiste à essayer de dégager des universaux du langage humain en s'appliquant à analyser les différences typologiques des langues. Ce type de linguistique peut en réalité se rattacher à la biologie, aux études cérébrales, car ce qu'elle prétend atteindre, ce sont nos structures mentales. À cet égard, on essaie au moyen de machines d'imiter ce que sont les structures mentales humaines et d'analyser la manière dont elles fonctionnent avec l'informatique, c'est-à-dire avec des robots humains. On n'a pas jusqu'ici atteint de résultats très probants.

*Secousse : Si on vous dit « littérature et ordinateur », à quoi pensez-vous d'abord, en tant que linguiste ?*

CH : Un grand nombre de gens qui ne lisent pas de livres peuvent avoir accès à des textes littéraires sur leurs écrans grâce à Internet. Voilà ma première idée. On peut avoir quelques scrupules à répondre de manière un peu intempestive à votre question, mais je crois que l'ordinateur n'est en rien, contrairement à ce qu'on a pu croire au début, un supplément hostile ou malfaisant pour la littérature.

*Secousse : Du point de vue du linguiste, qu'est-ce qui distingue la littérature des autres domaines d'utilisation du langage ? En quoi les ordinateurs peuvent-ils aider à mieux appréhender cette spécificité ?*

CH : La littérature est un domaine de l'activité du cerveau et du langage beaucoup plus codé et répondant à beaucoup plus de contraintes que les autres domaines car, lorsqu'on écrit, on n'adopte pas les mêmes normes que lorsqu'on parle – sauf dans les parties dialoguées des romans, où les dialogues imitent en principe la langue orale ; mais même là, il y a une intervention de l'auteur. La littérature est une activité capitale, universelle. Elle est définie par un style. Qu'est-ce que le style ? Une définition dit que c'est *l'art de s'exprimer par des moyens choisis*. En d'autres termes, quand vous vous exprimez, et même quand vous parlez, votre style, ce sont les moyens grammaticaux et lexicaux que vous avez choisis et qui vous caractérisent pour ceux qui vous connaissent, vous entendent, qui ont remarqué vos habitudes ou vos tics de langage.

Il faut ajouter, puisque j'ai insisté sur l'opposition entre l'écrit et l'oral, qu'il existe une littérature orale. Je suis un linguiste de terrain, j'ai fait beaucoup d'études un peu partout, de l'Afrique à l'Amazonie, en passant par la Chine et le monde arabe : partout, il y avait des gens qui n'avaient pas d'écriture (ou qui, ayant une écriture, n'étaient pas alphabétisés), mais qui avaient une littérature, que nous appelons *orale*. Elle conserve les traditions et les récits historiques, qui sont l'image et la conscience de peuples dans lesquels il n'y a pas d'écriture. Elle est en général proférée par des griots, en Afrique, ou par des réciteurs publics. C'est véritablement un *genre*, car la façon dont ils profèrent ces traditions et ces récits n'est pas du tout la même que lorsqu'ils parlent pour communiquer dans la vie ordinaire. On peut vraiment parler de *littérature orale* : comme toute littérature, elle est codée.

Quant à la question de savoir si les ordinateurs permettent de mieux appréhender la spécificité de la littérature, encore une fois, cela n'est vrai que dans la mesure où ils permettent d'analyser des collections énormes d'échantillons. Les ordinateurs, c'est essentiellement une aide au corpus. Les matériaux qu'on réunissait autrefois sous la forme de notes, d'enregistrements, de toutes sortes de moyens empiriques lents et peu puissants, peuvent aujourd'hui être réunis et commentés sur une très large échelle. Les ordinateurs ont considérablement facilité la constitution de corpus et l'accès à ces derniers.

*Secousse : Le processus de mémorisation d'un livre papier et d'un livre lu sur écran (livre numérique ou texte lu sur Internet) est-il le même ? Cette question fait-elle partie de vos interrogations ?*

CH : A priori, je ne pense pas que, sur ce plan, il y ait des différences fondamentales. La

seule différence est dans le temps que l'on met à revenir en arrière quand on veut mémoriser un passage qu'on a cru oublier : il suffit d'un clic sur l'ordinateur alors qu'avec un livre il faut manier les pages. Le mode d'accès purement visuel (dans le cas de la lecture sur écran), et à la fois visuel et tactile (dans le cas du livre), conduit-il à des différences dans l'appréhension d'un texte ? Certains comédiens disent que lorsqu'ils apprennent leur texte, ça leur est plus facile en utilisant un livre qu'un ordinateur. Pourquoi ? Parce qu'il y a un contact tactile avec le livre, et donc avec l'écrit : un contact presque charnel, qui n'existe pas avec un écran. La manipulation des pages d'un livre que l'on aime (c'est évidemment différent si l'on révise un cours dont on se fiche) a presque un aspect érotique, qui n'existe pas dans le cas de l'ordinateur.

À cet égard, nous sommes ramenés à une question très intéressante, mais qui relève moins de la linguistique ou de l'informatique que de la psychologie ou de l'étude des sens : la vision est-elle une aptitude plus ou moins intellectuelle que le toucher ? Le toucher est une aptitude que l'on pourrait considérer de prime abord comme extrêmement concrète mais, à la réflexion, ce n'est pas si évident. Pour beaucoup de gens (notamment les aveugles, pour qui le toucher permet la connaissance de ce qu'ils ne peuvent appréhender par la vision), le toucher est plus intellectuel qu'il ne le serait pour des voyants. Vous voyez qu'il est difficile d'apporter une réponse certaine à cette question.