

Le sanskrit et les mathématiques

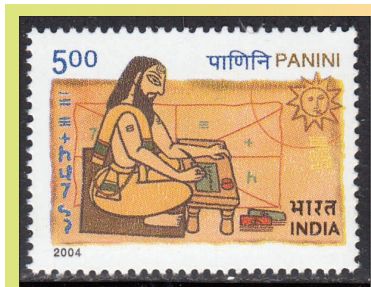
Gérard Huet

*Directeur de recherche émérite Inria
Membre de l'Académie des sciences*

Le sanskrit est la langue savante de la civilisation indienne. Ses plus anciens textes sacrés, constituant le Veda (ou littérature révélée), ont été transmis par une tradition orale remontant au deuxième millénaire avant notre ère. Cette transmission, très exacte, a été assurée par des procédés de récitation de formes combinatoires redondantes de ses hymnes. Les mètres très sophistiqués de ces poèmes montrent une maîtrise de l'analyse combinatoire sous la forme de la prosodie (*chandas*), érigée comme discipline de connaissance. La découverte du sanskrit par les orientalistes de la fin du XVIII^e siècle a été le point de départ de la linguistique comparative au XIX^e siècle, qui a engendré à son tour la linguistique générale occidentale, notamment grâce à la découverte des grammairiens antiques de l'Inde.

Des analogies avec les structures de données informatiques

Bien que transmis très fidèlement, le sanskrit védique est une langue archaïque complexe, et la compréhension des hymnes antiques est affaire de spécialistes. Par contre, le sanskrit classique est une langue essentiellement figée par la grammaire de Pāṇini, au IV^e siècle avant l'ère commune. Cette grammaire est un monument majeur parmi les œuvres intellectuelles de l'humanité, comme l'a déclaré le linguiste américain Leonard Bloomfield. Elle décrit avec une précision sans égale dans aucune autre langue les conditions sémantiques d'utilisation des procédés phonologiques, morphologiques, et structurels du langage pour obtenir des énonciations grammaticales.



Pāṇini (timbre indien émis en 2004).

© www.delcampe.net

lement correctes et signifiantes. De fait, elle n'a jamais été surpassée. Un commentaire magistral dû à Patañjali (II^e siècle avant l'ère commune) donne des éclaircissements et quelques compléments, et une refonte au XVII^e siècle réorganisa l'ordre des instructions sans en changer la teneur (*Siddhānta Kaumudī*).

Ce n'est que récemment que la grammaire de Pāṇini a été comprise à sa juste valeur. Ce texte, compréhensible seulement après un entraînement spécialisé, liste quatre mille formules très concises appelées *sūtras*.

Ce formulaire peut être compris comme le texte d'un programme informatique non déterministe permettant d'élaborer une énonciation à partir de l'intention de communication du locuteur. Le texte présuppose des connaissances phonétiques telles que la classification des phonèmes du sanskrit selon les traits phonétiques de position d'articulation, de distinction sourd-sonore, de présence ou absence de souffle, de nasalisation. La phonétique était reconnue de haute antiquité comme discipline de connaissance. La grammaire utilise certains phonèmes de la langue comme marqueurs métalinguistiques, anticipant de deux mille cinq cents ans les non-terminaux des grammaires de Chomsky ! Les phonèmes sont listés au début de la grammaire, entrelacés avec des marqueurs, afin de définir les sous-ensembles de phonèmes qui y interviennent. Ce codage est en fait une structure de données très concise pour dénoter ces sous-ensembles. Fait remarquable, il a été montré que ce codage est optimal. De fait, la concision de la grammaire fait appel à toutes sortes de techniques de partage et d'abréviation, ce qui lui permet d'être complètement mémorisable.

La grammaire de Pāṇini est le fondement d'une tradition de spéculation linguistique et sémiotique indienne, existant des siècles avant l'avènement, avec Ferdinand de Saussure au XIX^e siècle, de la discipline en Occident. Ce fait est bien reconnu par les linguistes, qui ont puisé largement à cette tradition, et une initiation au sanskrit est toujours de rigueur pour les étudiants en linguistique. En effet, la grammaire de Pāṇini définit les bases des calculs morpho-phonétiques modernes à base d'automates d'états finis, ainsi que celles de la syntaxe structurale à la Tesnière, qui a donné lieu aux développements récents des grammaires de dépendance, fondées sur des notions de rôles sémantiques d'agents du discours. D'autres disciplines intellectuelles de l'Inde antique, comme l'herméneutique (*mīmāṃsā*), ont discuté de notions d'économie, de cohérence et de pragmatique de la communication, qui n'ont été dégagées en Occident que dans les années 1960 avec la *théorie des actes de langage* de John Langshaw Austin et John Rogers Searle.

Il est maintenant reconnu que Pāṇini ne fut pas seulement un linguiste de génie, mais il fut un informaticien avant la lettre. En effet, sa grammaire, vue comme un programme informatique, repose sur des abstractions algorithmiques remarquables. En plus des structures de partage et de compression déjà mentionnées, on trouve en effet une méthodologie de réécriture algébrique sur les mots qui ne sera exposée en Occident qu'avec les systèmes de Post–Thue dans les années 1940. L'application d'une règle de réécriture dans un contexte utilise un codage d'enregistrements à l'aide de formules utilisant le paramètre morphologique de cas pour distinguer les champs, anticipant les spécifications de règles de transductions rationnelles mises au point en linguistique computationnelle dans les années 1980. De plus, les règles de la grammaire sont munies d'une notion de portée hiérarchique (*anuvṛtti*) qui anticipe la structure de contrôle d'appel par filtrage dans les langages de programmation modernes. Enfin, les méta-règles d'arbitrage du non-déterminisme définissent une structure de contrôle sophistiquée, anticipant la programmation orientée objet. On peut donc dire sans exagération que Pāṇini a posé les bases d'une méthode universelle de calcul relationnel bien avant les spéculations occidentales sur la notion de calculabilité au sein de la logique mathématique.

Une langue technique, idéale pour toutes les sciences

L'importance du sanskrit comme langue semi formalisée par l'existence d'une grammaire générative complète, dont le statut évolua de grammaire descriptive à celle de grammaire prescriptive, fit que cette langue standardisée et munie d'une sémantique rigoureuse joua dans une certaine mesure en Inde le rôle que jouèrent les mathématiques en Grèce comme support de l'argumentation rationnelle. La scholastique traditionnelle, utilisant le sanskrit muni d'une ontologie de notions lexicalisée, effectua un tournant méthodologique avec l'œuvre magistrale *Tattvacintāmaṇi* (« Joyau de la vérité ») du logicien Gaṅgeśa, fondateur de l'école de Mithilā de Nouvelle Logique (*Navyanyāya*) au XIV^e siècle. Cette école de sémiotique dialectique mit au point un calcul relationnel de notions formalisées en mots composés. Ces derniers (qui peuvent atteindre des dizaines de composantes) expriment de manière précise les relations entre objets du discours en explicitant leurs rôles sémantiques au sens de la grammaire du sanskrit et en levant les ambiguïtés possibles de la langue.

L'utilisation de ce langage formel (une sorte de méta-sanskrit) fut alors adoptée pour l'ensemble des sciences indiennes (*śāstra*). Cette langue technique, enseignée jusqu'à nos jours dans les établissements traditionnels d'enseignement, est indispensable pour comprendre tous les traités de la connaissance élaborés en Inde à partir du XIV^e siècle. Seulement connue en Occident d'une poignée de spécialistes, sa position



Représentation
supposée de Piṅgala.

[<https://bindass.org/aryabhata-is-not-zero-but-the-inventor-is-pingala/>]

précise par rapport aux formalismes mathématico-logiques modernes n'a pas encore été rigoureusement établie.

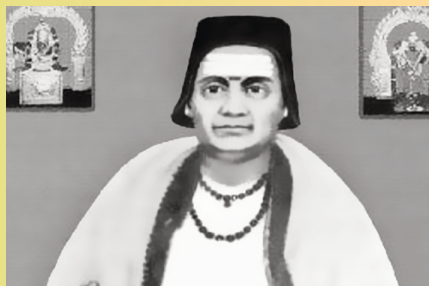
Les mathématiques ont connu un développement important dans l'Inde antique. L'analyse combinatoire fut développée pour rendre compte de la prosodie, dès l'ouvrage fondateur *Chandaśśāstra* de Piṅgala (antérieur à -400). Les considérations de métrique menèrent les théoriciens de la poésie aux premières investigations de la suite de Fibonacci et des cycles de de Bruijn. Une autre source de spéculations mathématiques était le problème de construire des autels de sacrifice obéissant à des conditions géométriques. Ainsi, Baudhāyana (IV^e siècle avant notre ère) donna l'approximation $577/408$ (soit 1,414215...) de la racine de 2, et énonça le théorème de Pythagore.

Dès les premiers siècles de notre ère, la notation décimale et le 0 étaient connus. L'astronomie développa pour ses besoins l'algèbre et la trigonométrie. Āryabhaṭa, astronome du Kerala (vers 500), évalua π à 3,1416 et l'année solaire à 365,3586805 jours; il proposa une théorie héliocentrique du système solaire, avec une Terre sphérique tournant sur son axe. Bhāskarācārya, qui dirigeait l'observatoire d'Ujjayinī, publia vers 1150 le *Siddhāntaśiromaṇi* («Couronnement de science»), somme de connaissances mathématiques et astronomiques en quatre chapitres: *Līlāvāṭī* (calcul arithmétique), *Bījagaṇita* (algèbre), *Grahagaṇitādhyāya* (calcul des planètes) et *Golādhyāya* (trigonométrie sphérique appliquée



Āryabhaṭa.

[<http://vedicciences.net/artiles/history-of-numbers-part-2.html>]



Bhāskarācārya.

[<http://economictimes.indiatimes.com/slideshows/people/indian-maths-wizards-and-their-ideas/bhaskara-ii-1114-1185/slideshow/40354408.cms>]

aux globes terrestre et céleste). Cette œuvre étonnante, rédigée en vers et pleine d'humour, présente le matériau mathématique sous une forme divertissante et donc facile à mémoriser. Il est émaillé d'applications calculatoires concrètes, tout en spéculant sur des objets abstraits comme les irrationnels et les limites. Le souci de la preuve est laissé aux commentateurs. Cette mathématique débridée est dans l'esprit artistique général de la littérature sanskrite, où les considérations esthétiques sont essentielles.

Des jeux de langues d'une richesse et d'une profondeur inouïes

Le drame sanskrit fut codifié de haute antiquité, et à sa suite la poésie fut l'objet de nombreuses spéculations esthétiques, développées dans un mouvement de critique littéraire impliquant des philosophes de premier plan comme Abhinavagupta. Les procédés rhétoriques, aussi bien phonétiques (assonance) que sémantiques (sens suggérés, sous-entendus), furent classifiés et critiqués. L'ambiguïté inhérente à la représentation continue des énonciations (*sandhi*) fut exploitée pour élaborer des récits se prêtant à un sens caché (*śleṣa*), ce qui donna naissance à un genre littéraire en soi, où un texte se prête à de multiples narrations simultanées. Ces jeux de langue, superposant par exemple à un panégyrique royal superlatif un sens caché injurieux, et où l'humour et l'érotisme voisinent avec le sacré, ont fait l'objet d'une poésie de cour extravagante auprès de laquelle les irrévérences du mouvement Oulipo paraissent des jeux d'enfant. On composa des poèmes se lisant comme l'histoire de Kṛṣṇa dans un sens, et l'histoire de Rāma en sens inverse (en lisant le poème à l'envers syllabe par syllabe). Un poème pouvait avoir jusqu'à sept interprétations différentes, en variant sa segmentation en mots. Ces jeux formels avec la langue présentent un continu avec la combinatoire mathématique, et ont forgé une tradition intellectuelle originale propre à la littérature sanskrite.

Le traitement informatique de la langue sanskrite a été entrepris depuis une quinzaine d'années. La grammaire de Pāṇini est progressivement traduite en termes de calculs effectifs, permettant de simuler la production d'énonciations conformes aux règles de la grammaire. Le problème inverse, visant à comprendre un texte sanskrit, est l'objet de recherches très actives. La compilation des règles morpho-phonétiques de Pāṇini en transducteurs rationnels inversibles a mené à la réalisation d'analyseurs lexicaux et étiqueteurs morphologiques utilisant des outils mathématiques intéressants, comme les *machines d'Eilenberg*, qui généralisent les automates d'états finis. L'utilisation d'algorithmes de résolution de contraintes sur des graphes de dépendance permet d'espérer disposer

d'ici peu d'analyseurs syntaxiques, utilisables pour annoter sémantiquement le corpus sanskrit, au moins pour la prose. Ceci ouvrira la voie à la manipulation par ordinateur du corpus de textes sanskrits, qui est considérable. Lorsqu'une quantité suffisante de textes auront été annotés par des linguistes et philologues avec l'aide de ces logiciels, on peut espérer que les techniques d'apprentissage profond permettront de mieux automatiser ces traitements, et d'avoir ainsi accès aux trésors de savoir de la langue sanskrite.

G. H.

Pour en savoir (un peu) plus :

Some features of the technical language of Navya-Nyāya. Sibajiban Bhattacharya. *Philosophy East and West* 40 (2), 1990.

Extreme Poetry. Yigal Bronner, Columbia University Press, 2010.

Pāṇini – A survey of research. George Cardona, Mouton, 1976.

Towards a formal regimentation of the Navya-Nyāya technical language. Jonardon Ganeri, in *Logic, Navya-Nyāya and Applications*, College Publications, 2008.

A Functional Toolkit for Morphological and Phonological Processing, Application to a Sanskrit Tagger. Gérard Huet, *Journal of Functional Programming* 15 (4), 2005.

Sanskrit signs and Pāṇinian scripts. Gérard Huet, in *Sanskrit and Computational Linguistics*, D.K. Publishers, 2016.

Materials for the study of Navyanyāya logic. Daniel Ingalls, Harvard University Press, 1951.

Śābdabodha and the problem of knowledge-representation in Sanskrit. Bimal Krishna Matilal, *Journal of Indian Philosophy* 16, 1988.

A Mathematical Analysis of Pāṇini's Śivasūtras. Wiebke Petersen, *Journal of Logic, Language and Information* 13 (4), 2004.

La grammaire de Pāṇini. Louis Renou, École Française d'Extrême-Orient, 1966.

Materials for the study of Navyanyāya logic. Daniel Ingalls, Harvard University Press, 1951.

Euclid and Pāṇini. Frits Staal, *Philosophy East and West* 15 (2), 1965.

The Aṣṭādhyāyī of Pāṇini. Rama Nath Sharma, Munshiram Manoharlal Publishers (six volumes), 1987–2003.