

# *La promotion de la traduction des langues vernaculaires en Inde passera-t-elle par le numérique ?*

of alternative francophone  
pour une francophonie en mode mineur

DOI : <https://doi.org/10.29173/af29519>



Bénédicte Diot

[benedicte.parvazahmad@inalco.fr](mailto:benedicte.parvazahmad@inalco.fr)

Institut National des Langues et Civilisations Orientales

**Résumé.** La politique linguistique de l'Inde, État plurilingue par excellence, se veut inclusive jusqu'aux langues les plus minoritaires avec une application multidirectionnelle de la traduction administrative protégée par la loi. En réalité cette politique est hiérarchisée. Elle met en avant l'hindi parmi les langues régionales et n'est pas appliquée pour de nombreuses minorités. Ceci transparaît dans les initiatives public/privé qui développent à grands pas depuis 2020 la localisation des technologies numériques pour les langues vernaculaires. La combinaison de traduction est le plus souvent de l'anglais vers les autres langues selon un ordre récurrent (hindi, puis les sept premières langues régionales, puis les autres langues répertoriées), indiquant la prééminence de l'anglais et la place symbolique des autres langues dans cette hiérarchie. Cette valorisation de la maîtrise des langues « du sommet » de la pyramide n'incite pas au développement de la traduction dans les langues minoritaires ou même entre « langues répertoriées » autres que l'hindi. De plus, la politique de traduction actuelle ne définit pas précisément les critères de sa mise en œuvre. Par ailleurs, depuis près de 10 ans, l'Inde a choisi de miser sur le développement de son économie par le numérique. Les internautes non anglophones constituent un marché gigantesque pour le secteur privé et un vivier de compétences pour peu qu'ils puissent accéder aux savoirs dispensés jusqu'ici en anglais. Ce développement vers les langues indiennes répond aussi à l'idéologie nationaliste prônée par le gouvernement en place. L'exemple de la diffusion de l'information lors de la pandémie du COVID-19 montre cependant que les langues des minorités les plus fragiles n'ont pas été prises en compte. Pourtant, les nouvelles technologies de l'information, en particulier l'IA, peuvent pourtant devenir des alliées dans la préservation et la traduction des langues minoritaires.

**Mots clés :** politique linguistique; Asie du Sud; numérique; traduction; langues peu dotées

**Abstract.** The language policy of India, a multilingual state *par excellence*, is intended to be inclusive, right down to the most minority languages, with a multidirectional application of administrative translation protected by law. In reality, this policy is hierarchical. It gives pride of place to Hindi among regional languages, and is not applied to many minorities. This is reflected in the public/private initiatives that have been developing digital localization technologies for vernacular languages at a rapid pace since 2020. The combination of translation is most often from English to other languages in a recurring order (Hindi, then the first seven regional languages, then the other listed languages), indicating the pre-eminence of English and the symbolic place of other languages in this hierarchy. This emphasis on mastery of the languages “at the top” of the pyramid does not encourage the development of translation into minority languages, or even between “listed languages” other than Hindi. What’s more, the current translation policy does not precisely define the criteria for its implementation. Moreover, for almost 10 years now, India has been banking on the development of its economy through digital technology. Non-English-speaking Internet users represent a gigantic market for the private sector, as well as a pool of skills, provided they can access the knowledge hitherto dispensed in English. This development towards Indian languages also reflects the nationalist ideology advocated by the current government. The example of the dissemination of information during the COVID-19 pandemic shows, however, that the languages of the most vulnerable minorities were not taken into account. And yet, new information technologies, particularly AI, can become allies in the preservation and translation of minority languages.

**Keywords:** language policy; South Asia; digitization; translation; low resourced languages

## INTRODUCTION

Le nombre des langues parlées en Inde comme langues maternelles<sup>1</sup> ou langues premières n'est pas exactement connu. Selon les critères gouvernementaux du recensement de la population qui a lieu tous les dix ans, il a ainsi varié de 1652 (1961) à un peu plus de 100 (1971), pour s'établir en 2011 à 122 langues reconnues parlées par des communautés d'au moins 10 000 locuteurs, les autres étant comptabilisées dans la catégorie hétéroclite « autres ». Mohanty propose une fourchette de 300 à 400 langues, tandis que le site d'Ethnologue en indique 456<sup>2</sup>. Quel que soit leur nombre, ces langues se répartissent en cinq grandes familles linguistiques et 40 % sont en danger, à des degrés divers<sup>3</sup>.

<sup>1</sup> Selon les termes du questionnaire passé par les recenseurs.

<sup>2</sup> Outre le critère du nombre de locuteurs, d'autres paramètres brouillent la classification : l'accès à certaines communautés rendant trop difficile la documentation de leur langue, les phénomènes de contact tellement anciens entre des langues rendant incertaine leur identification comme langue à part entière ou comme variante d'une autre langue, ou encore les déclarations sous des noms différents d'un même parler. Lors du recensement de 2001, les locuteurs étaient ainsi appelés à déclarer eux-mêmes leur langue maternelle, aboutissant à plus de 6 000 langues déclarées. Sur les 3 600 retenues, environ 2 000 étaient parlées par moins de 10 000 locuteurs et ont été répertoriées sous la rubrique « autres », ce qui laisse tout de même près de 1 600 langues.

<sup>3</sup> 197 langues indiennes sont en danger (UNESCO).

Pour satisfaire aux revendications linguistiques, le territoire de l'Union indienne a, depuis sa création en 1947, été plusieurs fois réorganisé, avec la création de nouveaux États<sup>4</sup> ou territoires associés à une langue majoritaire. Aujourd'hui, le pays est constitué en une fédération de 28 états et 8 territoires dont aucun n'est monolingue. Les 22 langues bénéficiant d'un statut dans l'annexe 8 de la Constitution et que nous désignerons dans la suite de cet article comme langues répertoriées, regroupent à elles seules 80 % de la population du pays, ce qui laisse quelque 280 millions de locuteurs parlant d'autres langues, dans des communautés regroupant jusqu'à plusieurs millions de locuteurs. Si la Constitution prévoit des garanties pour ces communautés linguistiques en minorité, en partie pour la traduction, elles sont sur le terrain difficilement applicables.

Par ailleurs, le récent essor d'internet, avec la baisse des coûts de communication, a entraîné le souhait d'une part importante de la population, non anglophone, d'accéder au numérique et d'en consommer les usages dans ses langues. Nous verrons si le virage du numérique pris par l'Inde contribue à mieux représenter les langues vernaculaires et les minorités linguistiques.

## SPÉCIFICITÉ DU MULTILINGUISME INDIEN

Le multilinguisme indien se caractérise par une grande fluidité entre les identités linguistiques et par sa structure organique dans laquelle l'on utilise de manière complémentaire différentes langues et l'on se passe souvent de la traduction.

Sans revenir sur l'histoire de la politique linguistique du pays, il convient de garder à l'esprit que le territoire, vaste et densément peuplé, est constitué de plusieurs grandes familles linguistiques (sans compter les isolats), que la langue numériquement la plus parlée, sans toutefois être majoritaire, est l'hindi-ourdou et qu'avant l'Indépendance en 1947, il n'y a pratiquement jamais eu de coïncidence entre la langue du pouvoir et celle du peuple sur un territoire donné. Comme en témoignent les œuvres littéraires depuis l'Antiquité, les langues ont une fonction sociale (Khushchandani, « Multilingual societies » ; Montaut) : l'emploi de chacune est circonscrit à un contexte situationnel privé ou public comme le cercle familial, le commerce, l'éducation, la religion ou l'administration.

On trouve trace de cette organisation dès les productions culturelles de l'Antiquité. Dans le théâtre sanskrit, on assignait aux personnages (rois, femmes, basses castes) des langues en fonction de leur place dans la société, et ces interactions n'étaient pas traduites lors des représentations.

Traditionnellement, la langue n'est donc pas conçue comme un moyen d'expression universel, chaque langue faisant partie d'une structure globale plurilingue, comme l'indique Srivastava cité par (Montaut 14). Le multilinguisme, qui prévaut chez les groupes minoritaires en cas de contact avec des langues dominantes, se maintient dans le temps, permettant la conservation de la langue avec des aménagements puisque les usages ne se font pas concurrence, mais se complètent (Mohanty ; Pandit). Les langues assurent des fonctions de communication complémentaires qui s'influencent (Bhatia et Ritchie 795) au

---

<sup>4</sup> Le dernier en date est l'État du Telangana en 2014.

sein d'une aire linguistique<sup>5</sup> commune.

Dans la zone hindiphone, qui rassemble le plus grand nombre de locuteurs, c'est l'hindi qui domine culturellement et administrativement : c'est la langue des grands médias, en particulier du cinéma. C'est aussi, au niveau de l'Union fédérale, la langue de l'administration centrale, dans sa variante enrichie de termes artificiels. En dehors de la zone hindiphone, les grandes langues régionales ou vernaculaires ont un rôle semblable au hindi dans un périmètre plus limité, celui de leur État. Les autres langues sont circonscrites à des usages privés, familiaux, religieux, dans tous les cas informels.

## MILLE-FEUILLE LINGUISTIQUE

Parallèlement à cette répartition complémentaire des langues, le multilinguisme indien se présente sous la forme d'une pyramide de prestige socioculturel. Mohanty parle de « double division » : d'un côté entre l'anglais et les grandes langues régionales, de l'autre entre ces langues reconnues, annexées à la Constitution, et les autres, dites minoritaires, bien qu'elles regroupent dans certains États davantage de locuteurs que la langue officielle. On trouve au sommet de cette pyramide l'anglais, suivi dans l'ordre par l'hindi, les autres langues régionales des États, les langues au statut officiel local, puis les langues de groupes plus restreints. Les strates les plus haut placées bénéficient de priviléges, bien que la Constitution et des lois linguistiques aient mis en place des garanties pour toutes les langues.

### *LE HAUT DE LA PYRAMIDE : ANGLAIS ET HINDI*

Selon la Constitution indienne, promulguée en 1950, l'Union indienne a pour langue officielle l'hindi (article 343), langue indo-aryenne, rattachée à la grande famille des langues indo-européennes. Cependant, cette langue n'est pas majoritaire, n'étant parlée que par 40 % de la population. Pour ménager les États du sud qui parlent des langues dravidiennes et représentent 20 % de la population, l'anglais a été conservé comme « co-langue officielle » (« co-official language ») alors qu'il n'est langue maternelle que de 0,3 % de la population et parlé par 3 à 11 % de celle-ci. L'anglais est la langue qui confère le plus de prestige. C'est la langue de l'élite, de la modernité, du savoir et du pouvoir. Sa maîtrise assure l'obtention des emplois les mieux rémunérés. C'est également dans cette langue que sont rédigés en tout premier lieu les textes de loi au niveau fédéral et qu'est rendue la justice au plus haut niveau. Autant de textes qu'il faudra traduire pour une population largement non anglophone.

Les textes législatifs confèrent, dans le domaine de l'administration fédérale, un statut à peu près équivalent au hindi et à l'anglais, avec une volonté du moins officiellement inscrite dans la Constitution de favoriser l'hindi. Les États n'utilisant pas l'hindi comme langue d'administration ont la possibilité d'utiliser au choix l'une ou l'autre langue dans leurs communications avec le pouvoir central ou les autres États (*Official Language Rules*, 1976).

Sur les plans législatif et judiciaire, l'anglais conserve sa prééminence : les textes de loi sont initialement

<sup>5</sup> Les phénomènes de contact montrent également une influence des langues minoritaires sur le développement des langues dominantes, comme le montre le contact entre le santali, langue austro-asiatique, sur le maithili, langue indo-aryenne proche de l'hindi.

rédigés en anglais; la Cour suprême utilise l'anglais et les minutes des procédures des hautes cours sont généralement en anglais, avec quelques extraits traduits (ce peut être les déclarations des justiciables, généralement en langue vernaculaire. Le terme « vernaculaire » désignera ici les langues indiennes en dehors de l'anglais, quelle que soit la taille de la communauté qui la parle).

En cohérence avec les objectifs de la Constitution, le Département de la langue officielle, créé en 1975 sous l'égide du ministère de l'Intérieur, vise à promouvoir l'usage de l'hindi dans tout le pays. Un comité de création terminologique, le Committee for Scientific and Technical Terminology (CSTT), est chargé de créer les néologismes en hindi pour combler le manque terminologique dans les sciences, la langue source de référence étant l'anglais. Ce comité normatif a donné naissance à une langue administrative très artificielle et incompréhensible par la masse de la population, bien loin de l'hindi parlé dans les situations informelles : son usage est donc limité, critiqué, voire rejeté.

On trouve ainsi déjà dans la strate de la langue vernaculaire la mieux représentée qu'est l'hindi des manques dans l'expression de plusieurs domaines et une diglossie entre la variété standard et les nombreuses variantes dialectales. Les autres langues vernaculaires d'administration, quant à elles, ne bénéficient pas d'autant de moyens de promotion.

### *LA STRATE INTERMÉDIAIRE : LES LANGUES OFFICIELLES*

En dessous de l'hindi se trouvent les langues vernaculaires officielles qui servent à l'administration locale de l'État et dans l'éducation. Les États de l'Union restent libres d'utiliser une autre langue officielle que l'hindi. Cet usage doit être entériné par la loi au niveau de l'État et s'accompagner du maintien de l'anglais comme alternative (article 345 de la Constitution). Dans la réalité, peu de fonctionnaires maîtrisent l'anglais au niveau le plus bas de la hiérarchie : la langue vernaculaire reste ainsi la plus utilisée et l'alternative en anglais purement fictive. La variété standardisée sous laquelle se présente la langue vernaculaire dans les services publics et l'éducation est assez éloignée de la version parlée par la masse des locuteurs, excluant ces derniers des pouvoirs public et économique.

En plus de ces dispositions, l'annexe 8 de la Constitution indienne présente une liste de 22 langues qui bénéficient de garanties et de moyens supplémentaires tels qu'une reconnaissance littéraire. Ce sont toutes des langues parlées par des communautés de plus d'un million de locuteurs, qui ont parfois une importance historique ou symbolique, comme le sanskrit ou le sindhi, langues officielles dans aucun État. À l'inverse, plusieurs langues officielles locales ne font pas partie de cette liste, comme par exemple le chhattisgarhi (18 millions de locuteurs) ou le mizo (830 846 locuteurs en 2011). On constatera dans la partie suivante que ces 22 langues ne sont dans les faits pas traitées à égalité.

### *LES LANGUES CO-OFFICIELLES*

Par ailleurs, si une part substantielle de la population souhaite utiliser une autre langue que la langue officielle de l'État, cela est possible en reconnaissant cette langue comme co-langue officielle dans tout ou partie de l'État ou du territoire (article 347 de la Constitution). Cette part doit être d'au moins 60 % de la population de la localité, mais un taux de 15 % doit suffire à faire publier les règlements dans une langue supplémentaire. Ainsi à Delhi coexistent quatre langues officielles : hindi, anglais, ourdou et pendjabi. L'article 348 assure que dans le domaine de la justice les plaintes et requêtes puissent être

rédigées dans une langue utilisée par l'Union ou l'État.

Ces dispositions sont généralement inappliquées sur le terrain, comme le montre l'exemple du Bihar où l'ourdou, deuxième langue officielle, est discriminée dans la législation locale<sup>6</sup> ), et les plaintes en ourdou sont souvent refusées par les services de police. Les représentations graphiques mêmes de l'ourdou tendent à disparaître de l'espace public, avec leur retrait des façades de bâtiments universitaires.

### *LA STRATE INFÉRIEURE : LES LANGUES DES MINORITÉS*

Le recensement de 2001 fait état de 3592 langues « maternelles » dont 1957 parlées par moins de 10 000 locuteurs. Ces minorités linguistiques sont en théorie protégées par la Constitution qui prévoit des dispositions pour protéger tout groupe revendiquant une langue, une écriture ou une culture qui lui est propre. Cela inclut un droit à ouvrir et gérer des établissements scolaires pour maintenir la langue, garanti par l'article 30. Le commissaire national aux minorités linguistiques, prévu par l'article 350-B, est chargé de veiller au respect de ces droits.

L'Union garantie également une instruction au niveau primaire dans les langues des minorités linguistiques (article 22, « Right of Children to Free and Compulsory Education Act—RTE », 2009), la disponibilité de manuels scolaires et d'enseignants, la possibilité de formuler des demandes à l'administration et de recevoir des demandes en ces langues. Dans les faits, ces garanties linguistiques ne sont pas appliquées par les États. Seules les communautés de plus d'un million de locuteurs<sup>7</sup> ont pu obtenir un enseignement primaire dans leur langue et le nombre de langues enseignées à l'école a diminué de moitié depuis 1970 (Mohanty). Les groupes numériquement inférieurs ne doivent donc s'attendre à ne recevoir aucun soutien. Les groupes de moins de 10 000 locuteurs, minuscules à l'échelle du pays, représentent 1 % de la population et sont généralement pauvres (Mohanty).

Le multilinguisme indien est donc spécifique en ce qu'il présente une « double fracture », d'une part entre une élite anglophone qui accapare le pouvoir et les masses non anglophones, et d'autre part entre ceux qui maîtrisent les grandes langues d'administration du territoire et les autres. Ce multilinguisme permet donc le maintien de langues minoritaires à côté de langues dominantes, mais il est inégal puisque certaines langues jouissent de priviléges en termes de pouvoir et de ressources et que les groupes les plus faibles sont contraints de s'adapter en réservant l'usage de leur langue à des domaines plus intimes et plus restreints, cette perte des domaines de compétence allant vers de l'affaiblissement de la langue. L'absence d'enseignement dans les langues des groupes minoritaires, parmi les plus précaires, engendre des difficultés dans l'intégration scolaire des enfants. Plongés dans le milieu linguistique de la langue majoritaire qu'ils maîtrisent mal, ils connaissent d'importants taux d'échec scolaire.

### **LA TRADUCTION VERS LES LANGUES VERNACULAIRES**

Intéressons-nous maintenant à la mise en pratique de la traduction dans le cadre de ce multilinguisme. L'administration l'encourage-t-elle ? A-t-elle les moyens d'appliquer les droits

<sup>6</sup> Voir la loi sur l'éducation *Bihar Education Service Rules* de 2014, où seul l'hindi est mentionné

<sup>7</sup> Soit 32 langues.

linguistiques promis par la législation fédérale ? Nous verrons qu'à défaut de moyens publics, c'est l'effet de masse des locuteurs des langues vernaculaires qui est en train d'assurer le développement de ces langues et de leur traduction, en particulier dans les médias.

## LES DISPOSITIONS SUR LA TRADUCTION DANS LES SERVICES PUBLICS

La Constitution et les lois linguistiques de 1963 et 1976 donnent un cadre sur les modalités de traduction dans les services administratifs sans en préciser les moyens. Comme nous l'avons vu précédemment, la véritable langue du pouvoir est l'anglais. C'est dans cette langue que sont rédigés les textes de loi, les jugements des juridictions supérieures et qu'ont lieu la plupart des débats au parlement.

Cependant, tout texte législatif fédéral produit en anglais doit être traduit en hindi puis dans les autres langues officielles. Les jugements et arrêts des hautes cours de justice, ou encore la correspondance en anglais entre les États doivent également être traduits vers l'hindi, mais pas dans les autres langues.

Dans le domaine judiciaire, l'Inde a adhéré au Pacte international relatif aux droits civils et politiques. À ce titre, un justiciable doit avoir l'assistance d'un interprète lorsqu'il ne comprend pas la langue de la procédure qui l'incrimine. Dans un pays où la majorité de la population s'exprime dans des langues vernaculaires tandis que les magistrats et les avocats sont formés en anglais, le besoin de traduction dans les tribunaux semble inévitable. Il n'existe pourtant aucun statut officiel de traducteur ou d'interprète en justice et l'intervention d'un interprète, généralement un fonctionnaire dont la maîtrise des langues est reconnue, se fait à discrétion du magistrat. Ce sont en général les avocats qui assurent une traduction officieuse à leurs clients. Les travaux de Berti sur le déroulement des audiences au tribunal montrent que les justiciables, qui ne s'expriment pas en anglais, n'ont pas la traduction des propos échangés dans cette langue entre les avocats, la formation de jugement et le greffier. De plus, les traductions des jugements de l'anglais vers les langues vernaculaires sont parcellaires.

Dans l'administration courante, le rapport de 2013 du Commissaire aux minorités linguistiques indique qu'aucun État n'appliquait les dispositions concernant la traduction des règlements et avis dans toute langue parlée par plus de 15 % de la population d'un canton (tehsil / taluka) ou d'un district, et d'ailleurs très peu d'États ont adressé des données au Commissaire, ce qui donne une idée de l'importance qu'ils accordent à la question.

C'est un fait, malgré les dispositions officielles, les langues n'ont pas toutes le même statut dans les institutions et peu bénéficient de traduction. Le besoin en la matière est immense en comparaison du personnel formé.

## LE TROPISME DE L'ANGLAIS

On constate d'autre part à travers toutes les dispositions de politique linguistique que le sens de la traduction va de l'anglais vers l'hindi en priorité, puis de l'anglais vers les autres grandes langues régionales. L'anglais est perçu comme la langue de référence, celle de la conceptualisation et du développement depuis l'époque coloniale (Montaut). C'est justement pour doter l'hindi des termes manquants dans les domaines scientifiques qu'a été créé le CSTT dans les années 1960 ou que travaillent les missions comme la National Translation Mission (NTM) et que sont proposées les formations à la

traduction en Inde. La création néologique se produit dans le sens de l'anglais vers l'hindi (voire vers une autre langue vernaculaire) et laisse apparaître de nombreux calques lexicaux ou syntaxiques dans les langues cibles. La traduction, littéraire ou non, est rare entre langues vernaculaires au point qu'une œuvre dans une langue vernaculaire A est souvent d'abord traduite en anglais avant d'être traduite dans une langue vernaculaire B. Ceci est d'autant plus dommage qu'il existe d'importants parallélismes syntaxiques ou lexicaux entre les langues vernaculaires de l'aire indienne.

## LES INITIATIVES DE TRADUCTION DES LANGUES VERNACULAIRES

Le gouvernement indien dispose de plusieurs institutions pour développer le rayonnement culturel des langues vernaculaires, promouvoir la traduction entre les langues vernaculaires et leur outillage informatique. D'autres projets gouvernementaux s'appuient sur des institutions nationales pour favoriser la traduction et la production dans les langues vernaculaires. On trouve aussi dans plusieurs universités des centres de recherche sur les langues en danger.

### *DÉVELOPPEMENT LINGUISTIQUE ET CULTUREL*

Comme on l'a vu, les traductions des langues vernaculaires vers l'anglais sont peu nombreuses, et elles le sont encore moins entre elles, prenant l'anglais comme pivot. Une des rares institutions à encourager cet effort dans le domaine littéraire est la Sahitya Akademi dont les actions concernent en priorité les 22 langues répertoriées et l'anglais, avec la remise chaque année d'un prix de traduction vers chacune de ces langues, ou l'organisation de rencontres plurilingues. Lorsque l'on examine la liste des textes sources traduits en 2019, on observe que l'hindi et l'anglais se taillent la part du lion : 7 sur 23 sont en hindi, 4 en anglais (soit déjà près de la moitié) et 3 en bengali, langues déjà largement favorisées par ailleurs. Aucune langue « *adivasi* » (bodo, santali) n'est représentée.

En faveur des langues « non répertoriées », l'Akademi propose également depuis près de 25 ans un prix intitulé « *Bhasha Samman* » qui récompense 3 ou 4 œuvres par an, pour lesquelles des langues en usage dans la littérature médiévale peuvent également concourir.

Cette institution a établi deux centres de traduction pour traduire des œuvres en langues vernaculaires vers l'anglais<sup>8</sup> et deux centres de littérature orale pour les langues tribales (Centres for Tribal and Oral Literature) lui sont rattachés, le NECOL et le COTLIT, collectent les littératures orales de ces langues et les traduisent vers des langues vernaculaires répertoriées.

D'autre part, plusieurs universités, comme Shantiniketan (Vishva Bharati) au Bengale occidental, l'Université dravidienne ou l'Université du Kerala dans le sud, collectent des données sur les langues en danger et mettent à disposition du public des ressources lexicales (CFEL). Leur rôle reste pour l'instant plus documentaire que productif de contenus élaborés.

### *TRADUCTION TECHNIQUE : LE CSTT ET LA NATIONAL TRANSLATION MISSION (NTM)*

<sup>8</sup> Le site internet de l'Akademi n'enregistre toutefois une activité qu'entre 2001 et 2003.

La problématique de la création terminologique évoquée précédemment pour l'hindi vaut également pour les autres langues régionales. Le CSTT est chargé, dans une moindre mesure que pour l'hindi, d'assurer cette tâche et publie des lexiques et dictionnaires.

D'autre part, la National Translation Mission, créée autour de l'année 2000, a pour objectif de rendre les savoirs des grandes disciplines disponibles au plus grand nombre en les traduisant dans les grandes langues vernaculaires. Il existe en effet un fossé entre d'une part une élite urbaine qui dispose d'un corpus de textes scientifiques nationaux et internationaux en anglais et d'autre part une masse d'étudiants qui n'y ont pas accès de par leur absence de maîtrise de l'anglais. Il leur sera plus facile d'absorber ces sujets dans leur langue maternelle plutôt que de devoir se faire expliquer les textes en anglais, comme c'est généralement le cas.

L'institut traduit, forme des traducteurs et développe la traduction assistée par ordinateur. Le projet NTM dépend du Central Institute of Indian Languages (CIIL) qui produit des méthodes d'apprentissage des langues vernaculaires, forme des enseignants à leur pratique et mène des recensements linguistiques. Les ambitions du CIIL sont importantes dans le domaine du numérique où les institutions publiques ont pris du retard : des projets de collecte et d'analyses de corpus, qui permettront d'alimenter la TAO et le développement informatique des langues via le traitement automatique des langues (TAL), voient à peine le jour et ne sont pas encore disponibles pour le grand public (site en construction en 2021).

#### *LE LD-CIL (LINGUISTIC DATA CONSORTIUM FOR INDIAN LANGUAGES)*

Ce projet, également hébergé par le CIIL, porte sur le volet TAL des langues indiennes : il traite des corpus numériques afin de développer la linguistique outillée et la traduction assistée ou automatisée dans les langues vernaculaires.

Le centre dispose de corpus en 24 langues, dont l'anglais. Le plus important est le corpus hindi, avec 30 millions de tokens, suivi par le tamoul, avec un corpus trois fois moindre.

Le pays se dote de divers outils normatifs pour la traduction et le développement des langues vernaculaires. Il a pris le virage du numérique tardivement. Or aujourd'hui si une langue entend assurer sa survie, elle doit être représentée sur Internet, ce qui sous-entend qu'il existe des outils développés en ces langues pour produire du contenu numérique et le rendre accessible à la communauté des locuteurs.

### **L'ESSOR DES MÉDIAS ET DU NUMÉRIQUE : UNE REPRÉSENTATION ACCRUE DES LANGUES VERNACULAIRES**

À la fin des années 1990 et au début des années 2000, les technologies numériques étaient perçues comme renforçant la fracture entre une élite anglophone et le reste de la société non anglophone, car les outils et les formations pour les maîtriser n'existaient qu'en anglais. Cette perception a été balayée depuis quelques années avec la démocratisation de l'accès à internet et la volonté des énormes communautés non anglophones d'accéder aux mêmes médias que l'élite anglophone. Ces technologies apparaissent à présent comme un moyen de mieux représenter les langues vernaculaires. L'administration

indienne, au niveau fédéral ou central, a pris le virage du numérique à la sortie de la crise sanitaire du COVID-19. Elle ne remplit toutefois pas encore les objectifs escomptés, tandis que les sociétés des technologies de l'information cherchent à capter ce nouveau public non anglophone pour se développer. Nous observerons ces développements à travers deux cas concrets, celui de la traduction lors du début de la pandémie de COVID-19 et le développement de réseaux sociaux locaux dans les langues vernaculaires.

## LA VERNACULARISATION DES MÉDIAS INDIENS

Tout commence avec la libéralisation des médias audiovisuels indiens en 1992 qui introduisent de nouveaux modèles de consommation. Les premières chaînes privées sont en anglais, puis rapidement se développent des distributeurs locaux de télévision câblée diffusant plus ou moins légalement des films populaires dans les langues vernaculaires. Le paysage audiovisuel se structure petit à petit avec l'arrivée de chaînes généralistes et d'infos en continu dans les langues vernaculaires. Ces chaînes proposent une offre plus variée et plus moderne que les variantes en langues locales de *Doordarshan*, la télévision d'État.

La vernacularisation du paysage médiatique indien va se poursuivre avec l'essor du numérique. Dans un premier temps, on relaie la presse écrite à travers des sites internet. Ces sites permettent à la diaspora indienne, la première équipée en accès internet, de garder le lien avec ce qui se passe dans la région d'origine. Le site NRIOL (Non Resident Indians On Line), qui s'adresse à la diaspora, recense plusieurs dizaines de journaux en ligne dans une douzaine de langues vernaculaires.

Le développement du numérique dans les langues locales s'accélère avec la baisse des coûts d'internet autour de 2017. Les Indiens vont s'équiper progressivement et développer leurs propres contenus. On trouve ainsi aujourd'hui 22 langues indiennes représentées sur *Wikipédia*, dont le bihari-bhojpuri, le tulu et l'avadhi qui ne sont pas des langues répertoriées. La représentation des langues indiennes reste cependant modeste sur ce média en comparaison du poids démographique du pays : on peut y lire en 2024 seulement 163 000 articles en hindi (le basque en propose trois fois plus), souvent moins longs que les articles des langues dominant l'espace du net.

## RÉVOLUTION NUMÉRIQUE ET LOCALISATION

Deux types d'acteurs s'engagent dans le tournant du numérique en Inde : le gouvernement, qui souhaite automatiser les traductions et les échanges entre langues indiennes dans l'éducation et l'administration, et les sociétés des technologies de l'information. Tous deux sont liés dans des partenariats public-privé.

### LE VIRAGE NUMÉRIQUE DU SECTEUR PUBLIC

L'usage d'Internet a débuté en Inde il y a moins de 30 ans. Cette dernière décennie, son accès s'est démocratisé avec l'effondrement du coût des communications et des téléphones intelligents (Chauhan Abhishek et al. 70), si bien que le pays se situe en deuxième position mondiale en 2024 quant au nombre d'utilisateurs, représentant un marché de 833 millions d'individus (pour 530 millions en 2018), malgré un taux de pénétration faible de 52 %. L'accès au net n'est plus réservé à une élite

anglicisée, dont le poids est bien inférieur.

Dans sa National Education Policy (NEP 2020), le gouvernement reconnaît le retard pris par le pays dans la préservation des langues vernaculaires, qu'il s'agisse de terminologie ou de néologie, de documentation ou simplement de la possibilité de leur pratique. Il déclare souhaiter promouvoir les enseignements supérieurs en langues régionales, créer un institut et des diplômes supérieurs de traduction et d'interprétation (l'IITI, Indian Institute of Translation and Interpretation), et faciliter la collecte collaborative de corpus pour les langues en danger. Ce texte d'intention permet le lancement par la suite de chantiers encore en cours qui font appel à l'intelligence artificielle (traduction automatique, synthèse et reconnaissance vocale, etc.).

Le projet *Udaan* en 2023, mis en place conjointement par l'IIT Bombay et le gouvernement du Maharashtra a pour but de traduire les livres scientifiques en anglais vers des langues indiennes, partant du principe que l'absence de formation dans les langues premières est un frein pour l'économie indienne, comme le présente la page d'accueil du site internet :

The lack of native language instruction also has a significant economic impact. Of the top twenty countries in GDP per capita, all of them use the language of the common people for higher and technical education. Most of these countries do not use English. In the bottom twenty economies, most of the countries are those which persist with a colonial language, not the mother tongue of the majority of people. The transition to mother tongue education requires significant investment in language infrastructure. » (Udaan Project)

Cette théorie est aussi soutenue par des auteurs et universitaires comme Sankrant Sanu dans son livre *The English Medium Myth*. Les fonctionnalités de traduction de documents n'ont pas pu être testées par l'auteure et la fonctionnalité de traduction instantanée n'est pas encore complètement déployée. Sous l'égide de la NTM, le projet *Digital India Bhashini*, démarré en 2022, est une plateforme de traduction utilisant l'intelligence artificielle (IA). Il met l'emphasis sur l'utilisation de l'oral, à laquelle aspirent la majorité des utilisateurs des langues vernaculaires. On y trouve la fonctionnalité de reconnaissance vocale, qui permet de transcrire un discours oral, ou de synthèse vocale, pour oraliser des textes. C'est le système qui a été utilisé pendant la campagne électorale de 2024 pour traduire les discours du candidat sortant Narendra Modi à l'oral en 13 langues régionales. La plateforme existe sous forme d'application pour les portables.

Enfin, à l'intérieur du projet *Bhashini* se trouve le projet collaboratif *Bhasha Daan* (« don de langue ») de collecte de corpus dans des langues indiennes. Quatre tâches sont proposées sous forme ludique : on peut transcrire dans sa langue ce que l'on entend, décrire ce que l'on voit, s'enregistrer à l'oral ou traduire des textes écrits. 23 langues sont pour l'instant représentées. L'objectif est d'apporter une contribution humaine aux modèles utilisés par l'IA, permettant de corriger les algorithmes.

## LA PART DU SECTEUR PRIVÉ

Au-delà des usages éducatifs, les habitants des villes moyennes, des zones rurales<sup>9</sup> ainsi que les couches de la population non anglophones veulent utiliser les réseaux sociaux dans une langue qui leur permette de s'exprimer le plus intuitivement possible, en particulier à l'oral. La catégorie des « vernacular

---

<sup>9</sup> 92 % des ruraux utilisent le téléphone portable pour naviguer sur internet, d'après Media Monitor Ownership India.

users » constitue un marché gigantesque qui augmente plus rapidement que celle des utilisateurs anglophones<sup>10</sup> si bien qu'il représente 95 % des utilisateurs de *YouTube* en 2022 (Aggarwal).

Cette manne ne peut échapper aux grands acteurs commerciaux qui doivent faciliter l'accès du consommateur à leurs produits. De nombreuses jeunes pousses indiennes sont créées et travaillent sur les projets gouvernementaux. D'autres sont rachetées par les grands groupes du secteur ou disparaissent. Le développement de la localisation dans les langues les plus importantes en nombre de locuteurs (hindi bien sûr, suivi du bengali, du tamoul, du télougou et du marathi) est rapide, facilité par la diffusion de normes d'encodage international des divers systèmes d'écritures indiennes. Les normes locales ISCII sont ainsi remplacées par les normes Unicode.

## TRADUCTION SUR L'ESPACE NUMÉRIQUE : LE CAS DU COVID-19

L'administration indienne, qui a opéré ces dernières années un virage numérique conséquent avec la mise en ligne de nombreux services publics, a longtemps peiné à accompagner les aspirations d'un public non anglophone ou non hindophone à accéder aux informations. Le gouvernement affirme sa position nationaliste en soutenant quelques initiatives de développement de réseaux sociaux et d'applications alternatifs aux grands groupes, mais en cas de crise, ce sont les initiatives des ONG et de groupes privés qui assurent la transmission des informations aux communautés linguistiques non représentées.

La volonté de recourir de plus en plus au numérique dans l'administration se matérialise par le développement d'un portail d'information en ligne pour centraliser les services. Ici encore, on constate un effort de traduction vers l'hindi, avec une accessibilité à pratiquement tous les contenus dans cette langue, ce qui n'était toujours pas le cas il y a quelques années encore. En revanche, il n'est possible d'accéder à un contenu partiellement traduit que dans 11 langues répertoriées sur 22. Les langues représentées sont celles qui comptent les plus fortes communautés de locuteurs, à l'exception de l'ourdou, pratiquement jamais présent sur les sites gouvernementaux. La plupart des sites des États proposent, pour leur part, des contenus entièrement en anglais. Or la limitation ressentie par les Indiens qui n'utilisent pas l'anglais sur internet est bien réelle : 60 % d'entre eux ne trouvent pas de contenus dans leurs langues ni de « support langue », ce pourcentage montant à 94 % pour l'accès à l'information générale (KPMG/Google).

La prise en compte de la traduction vers les langues vernaculaires, qu'elles soient majoritaires ou minoritaires, est donc très parcellaire dans l'administration centrale ou régionale. Or s'il est un événement qui a rappelé l'importance de la traduction dans ces langues, c'est bien la crise sanitaire du COVID-19. Dès le début de l'épidémie, le besoin rapide de traduction a été très fort partout dans le monde, autant pour diffuser les consignes de prévention que pour dissiper les rumeurs. En Inde, les initiatives de traduction officielle ont, encore une fois, privilégié l'hindi puis les grandes langues régionales habituelles. On peut ainsi suivre en ligne l'évolution de la pandémie en hindi et partiellement dans les 11 autres langues déjà mentionnées<sup>11</sup>.

Il faut rechercher des initiatives non gouvernementales, et même en dehors de l'Inde, pour obtenir des informations traduites dans des langues minoritaires. L'initiative public-privé des GAFAM et

<sup>10</sup> 18 % contre 3 % entre 2016 et 2021 (KPMG/Google).

<sup>11</sup> Le corps de la page est traduit, mais les documents indexés par des hyperliens sont tous en anglais.

d'universités américaines, TICO-19 (Anastopoulos et al.), reprend les informations fournies par l'OMS. Au Japon, l'International Christian University a coordonné la traduction dans plus de 125 langues (« COVID-19 Myth Busters in World Languages ») — dont 17 langues indiennes minoritaires — de documentation sur la maladie, en lien avec des universitaires indiens concernés par les langues en danger<sup>12</sup>. Sur *Facebook*, le groupe Adivasi Lives Matter a contribué à la traduction en ho, népali et kodava dès le début de la pandémie. L'initiative *Endangered Language Project*, menée par des universités américaines, a proposé une plateforme intitulée « COVID-19 Information in Indigenous, Endangered, and Under-Resourced Languages » regroupant les ressources écrites ou vidéo d'autres ONG, dont Adivasi Lives Matter et VirALLanguages<sup>13</sup>, allant de la simple affiche avec les consignes sanitaires, aux vidéos et brochures pour lutter contre la désinformation. 61 langues indiennes y sont documentées, dont 46 non répertoriées<sup>14</sup>. On y trouve peu de sources officielles indiennes originales, ce qui confirme le manque d'initiative de traduction par le gouvernement à destination de communautés pourtant plus vulnérables que la moyenne.

## CRÉATION D'APPLICATIONS EN LANGUES VERNACULAIRES : UN OUTIL D'ÉMANCIPATION ?

Les nouveaux internautes non anglophones passent progressivement d'un usage passif à un usage actif de leurs langues : à la simple traduction de pages web ou la numérisation de documents en langues vernaculaires s'ajoutent la construction de pages web et le développement d'applications imitant des réseaux sociaux déjà existants, en les adaptant aux usages locaux. Ces applications fleurissent et se pérennisent avec plus ou moins de succès.

Si *Whatsapp* (Meta) reste très utilisée à travers le pays et a réalisé des efforts de localisation, les applications en langues vernaculaires « made in India » comme *ShareChat*, *Moj*, *Koo*<sup>15</sup>, *Roposo* ou *Chingari*, sont plébiscitées (blog de Chingari, Startup Talky, News 18) par les nouveaux utilisateurs qui peuvent enfin s'exprimer et lire dans leur langue première. Certaines, comme *Tooter*, mettent en avant l'aspect nationaliste de la démarche, on trouve en effet sur *ShareChat*, alternative à *Instagram*, plusieurs langues d'Inde du Nord n'appartenant pas à la liste des langues répertoriées : le bhojpuri, le haryanvi et le rajasthani. Cela traduit le besoin des millions de locuteurs qui les parlent d'échanger plus spontanément que dans une langue standardisée. Leur poids démographique suffit à faire évoluer le marché (Youth Ki Awaz). Pour les acteurs du numérique, le vecteur de la langue vernaculaire permet d'obtenir plus facilement de l'utilisateur des réponses aux stimuli proposés.

Conscient des enjeux stratégiques commerciaux et géopolitiques, le gouvernement indien soutient ces

<sup>12</sup> En particulier dans les Centre for Endangered Languages of Kerala (Celk) et Centre for Preservation of Endangered Dravidian Languages (CPEDL).

<sup>13</sup> Sous l'égide de l'OMS, VirALLanguages s'attache aux langues de cinq pays dont le Pakistan, qui partage des langues avec l'Inde.

<sup>14</sup> Les langues représentées sont les suivantes : ahom, anami, angika, assamais, adiya, balti, bengali, bhojpuri, bodo, burushaski, chakma, dimasa, bhotia1, gaddi, gujri, goujarati, lai, ho, irula, kannada, karbi, cachemiri, kodava, kok-borok, kolami, konkani, koya, kurumba, lepcha, liangmai, limbu, maithili, malayalam, marathi, meitei, mizo, muduga, muthuva, népali, oriya, ollari gadaba, pahari, paniya, pendjabi, rajbongshi, santali, mey, shina, sindhi, lambadi, sunu-war, khamti tai, tamang, tamoul, tangkhul, télougou, tutsa, ourdou, wancho, warli et zou.

<sup>15</sup> *Koo*, version indienne de *X* (*Twitter*), a fermé début juillet 2024.

entreprises indiennes dont il met en avant le caractère local, qu'il s'agisse de jeunes pousses ou d'importants groupes d'investissements comme Reliance. Il interdit au même moment à des concurrents étrangers l'accès au marché, comme pour *TikTok* depuis 2020. Cette démarche fait écho à la volonté affichée de « décoloniser » le numérique indien inscrite dans sa stratégie *Atmanirbhar Bharat* (« Inde autosuffisante » ou « confiante en soi »), même si les capitaux investis appartiennent paradoxalement en grande partie aux GAFAMs.

Lorsqu'on observe de plus près le contenu de ces applications, on constate toutefois que la créativité propre à chaque langue est très limitée. Sur l'application *Chingari*, par exemple, les versions en bhojpouri ou en rajasthani sont en réalité en hindi standard et les contenus (de courts montages vidéo de chansons ou de films) empruntent tous au même répertoire des films hindi standards. *ShareChat*, en revanche, présente des variantes de localisation. Les contenus localisés empruntent massivement à l'anglais et ne sont souvent que la transcription de cette langue dans d'autres systèmes d'écriture (par ex. « account », « post », « content », « link »).

Les outils numériques se développent ainsi dans des domaines plus variés qu'autrefois, permettant à de nouveaux publics de participer à la vie des réseaux numériques, ce dont il faut se réjouir. Cependant, plusieurs aspects viennent tempérer ce constat. D'une part, la variété de langue utilisée dans les vernaculaires emprunte ou calque beaucoup à partir de l'anglais qui conserve son statut dominant. De nombreux termes technologiques sont simplement transcrits dans les écritures régionales, ce qui ne développe pas la créativité terminologique. Parfois, l'utilisation du nom d'une langue de diffusion plus locale semble purement cosmétique, sa localisation étant identique à la langue standard. D'autre part, ce développement se limite à une dizaine de langues, dont le poids des communautés de locuteurs — et de consommateurs potentiels — ne peut être ignoré par les acteurs économiques. Enfin, cet essor ne concerne pas les langues très minoritaires. Pour celles-ci, l'accès à l'information reste problématique et repose sur des initiatives extérieures à l'État.

## CONCLUSION

À travers ce survol de la situation linguistique des différentes communautés de locuteurs en Inde, on constate que si les droits linguistiques sont attachés en théorie à toutes les communautés, quelle que soit leur taille, dans les faits les communautés les plus nombreuses ont la capacité de peser dans la prise en compte des besoins de traduction et d'accès à l'information, car elles représentent un poids économique et politique fort. La démocratisation du numérique les a rendues visibles et leur a permis, grâce à des acteurs privés, de se développer dans des domaines jusqu'ici réservés à l'anglais (communication et divertissement, technologies du numérique). Le gouvernement accompagne timidement ce mouvement à travers la traduction administrative qui reste limitée à une dizaine de grandes langues. La traduction pragmatique officielle est très lacunaire pour la douzaine de langues répertoriées restantes. Dans les cas où cette traduction existe, elle est très influencée par l'anglais et la créativité traductionnelle demeure faible. Quant aux langues plus minoritaires, parlées par des communautés précaires et dispersées, elles n'ont pratiquement aucune visibilité, comme l'a montrée l'absence de traduction officielle lors de la crise du COVID-19.

Les nouvelles technologies de traitement automatique des langues (TAL) comme les modèles d'apprentissage machine non supervisés ont connu ces dernières années des avancées importantes. Ces modèles de langue indépendants des données linguistiques permettent de développer des contenus à partir

de corpus oraux comme écrits fournis comme paramètres. Leur mise en œuvre nécessite de moins en moins de connaissances spécialisées ou de ressources, ce qui les rend faciles à adapter aux communautés visées et aux données orales. Des modèles sont déjà en accès libre comme sur *Bhashini* pour la synthèse vocale (Gokul et al.) ou le modèle VITS de Coqui-AI qui a été utilisé pour la reconnaissance vocale du *Chatisgarhi*. Ce type de technologie permettrait de monter des projets rapides et peu coûteux pour que l'accès à l'information et la communication dans les langues peu dotées soient assurés.

## BIBLIOGRAPHIE

- Berti, Daniela. *Regimes of Legality. Ethnography of Criminal Cases in South Asia*. OUP, 2015.
- Bhatia, Tej K. et William C. Ritchie. « Bilingualism in South Asia. » *The Handbook of Bilingualism*, 2004, pp. 780–807.
- Center for Endangered Languages (CFEL), *Progress Report 2014–2018*, Visva-Bharati, Santiniketan, 2018.
- Gokul, Karthik Kumar et al. « Towards Building Text-to-Speech Systems for the Next Billion Users. » ICASP, 2023, <https://arxiv.org/pdf/2211.09536.pdf>.
- Gouvernement indien. *Official Languages Rules*, 1976.
- Khubchandani, Lachman M. « Languages of Instruction in Plurilingual India. » *International Review of Education*, vol. 24, no. 3, 1978, pp. 75–80.
- Khubchandani, Lachman M., « Multilingual societies: Issues of identity and communication. » *Sociolinguistics*, vol. 16, no 1, 1986, pp. 20-34.
- Mohanty, Ajit K. « Languages, inequality and marginalization: implications of the double divide in Indian multilingualism. » *International Journal of Sociology of Languages*, no. 205, 2010, pp. 131-154.
- Montaut, Annie. « L'anglais en Inde et la place de l'élite dans le projet national. » *Hérodote*, vol. 4, no. 115, 2004, pp. 63-89.
- Pandit, Prabodh Bechardas. *Language in a plural society: the case of India*. Channa Memorial Committee, 1977.
- Sanu, Sankrant. *The English Medium Myth: Dismantling Barriers to India's Growth*. Garuda Prakashan Private Limited, 2018.
- UNESCO. *Atlas of the World's Languages in Danger*, 2010.

## WEBOGRAPHIE

- « Adivasi Lives Matter, Safety Against COVID-19. » *Facebook*, Towards Building Text-to-Speech Systems For the Next Billion Users.[https://www.facebook.com/AdivasiLivesMatter/posts/827595691049916?\\_tn\\_=R](https://www.facebook.com/AdivasiLivesMatter/posts/827595691049916?_tn_=R).
- Aggarwal, Ankit. *Road To 2022: The Increasing Appetite for Vernacular Content*. Social Samosa, 2022, <https://www.socialsamosa.com/2022/01/road-to-2022-increasing-appetite-vernacular-content/>.
- Akademi, Sahitya. *Annual Report*. <http://sahitya-akademi.gov.in>.

- Anand, Sarika et Bajj, Apoorva. *Koo: The Rise and Shutdown of India's Homegrown Microblogging Platform*. Startup Talky.com, <https://startuptalky.com/koo-success-story/>.
- Anastopoulos et al. « TICO-19: the Translation Initiative for COVID-19. » Carnegie Mellon University, 2020, <https://arxiv.org/pdf/2007.01788.pdf>.
- An Informed Citizen. *Tapping The Power Of Language: Vernacular Media Apps Are On The Rise. Youth Ki Awaz*, 2019, <https://www.youthkiawaz.com/2019/08/tapping-the-language-power-vernacular-media-apps-and-their-issues/>.
- Abhishek, Chauhan et al. *Vernacular is NOW, not the future. Redseer*, 2019, <https://redseer.com/reports/vernacular-is-now-not-the-future-a-300-bn-opportunity-today/>.
- Bhasha Daan. Ministry of Electronics and Information Technology, National Language Translation Mission, Bhasha Daan, <https://bhashini.gov.in/bhashadaan/en/home>.
- Commission for Scientific and Technical Terminology. *Shabd*, <https://shabd.education.gov.in/>.
- COVID-19 Myth Busters in World Languages, 2022, <https://covid-no-mb.org/>.
- Das Shouvik. *The Next Billion: 25 Years In, Vernacular India to Give Our Internet a New Path. News 18*, 2020, <https://www.news18.com/news/tech/the-next-billion-25-years-in-vernacular-india-to-give-our-internet-a-new-path-2783649.html>.
- Data Leads. « Media Consumption. » *Media Monitor Ownership India*, 2019, <https://india.mom-rsf.org/en/context/media-consumption/>.
- Data Peortal, *Digital 2024: India*, [datareportal.com/reports/digital-2024-indiac](http://datareportal.com/reports/digital-2024-indiac).
- Devnagri, <https://devnagri.com/>.
- Eberhard, David M., Gary F. Simons et Charles D. Fennig, éditeurs. *Ethnologue : Languages of the World*. Twenty-fifth edition. SIL International, 2022. <https://www.ethnologue.com/guides/countries-most-languages/>.
- IIT Bombay. *Udaan project*, 2022, <https://udaanproject.org/AboutUs>.
- KPMG/Google. *Indian Languages—Defining India's Internet*. avril 2017, [https://kpmg.com/ky/en/home/insights\\_new/2017/04/indian-language-internet-users.html](https://kpmg.com/ky/en/home/insights_new/2017/04/indian-language-internet-users.html).
- Ministry of Education, Linguistic Data Consortium for Indian Languages. *Datasets (Linguistic Resources)*. LDC-IL, <https://www.ldcil.org/>.
- Ministry of Electronics and Information Technology, National Language Translation Mission. *Digital India Bhashini*. <https://bhashini.gov.in/>.

National Science Foundation: Endangered Language Project. *COVID-19 Information in Indigenous, Endangered, and Under-Resourced Languages*, <https://endangeredlanguagesproject.github.io/COVID-19/>.

Organisation Mondiale de la Santé. *VirALLanguages*, 2020, <https://virallanguages.org/fr/accueil/OMS>.

ShareChat. *About*, <https://sharechat.com/about>.

Vernacular Newsline—Indian Regional Newspapers. *Non Resident Indians Online*, <https://www.nriol.com/news/newsline/vernacular.asp>.

« What Is Chingari And Why Is It So Popular In India? » *Chingari*, 2020, <https://blog.chingari.io/what-is-chingari-and-why-is-it-so-popular-in-india/>.

Wikipedia. *List of Wikipedias*, [https://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](https://meta.wikimedia.org/wiki/List_of_Wikipedias).