

Les langues indo-aryennes (à paraître dans L'Encyclopédie des Sciences du langage) Annie Montaut

1. Contexte

Précisons d'emblée que le terme « aryen » est utilisé ici par commodité : cette étiquette traditionnelle ne recouvre rien de linguistique ni d'ethnique, mais une qualification de type social et religieux¹.

Les langues IA modernes, toutes issues du sanscrit et de sa prakritisations, regroupent aujourd'hui près d'un milliard de locuteurs en Asie du Sud, sans compter la diaspora. Six fonctionnent comme langue nationale : le hindi en Inde, (42% des locuteurs Indiens, soit 400 millions en incluant les 'dialectes'), mais aussi aux Iles Fiji et Maurice où il est langue officielle associée, le bengali au Bangladesh (100 millions), l'ourdou au Pakistan (8 millions), le népali au Népal (10 millions), le singhalais à Sri Lanka (env. 13 millions), le dhivehi ou maldivien dans la République Maldivie. Treize ont un statut officiel dans un ou plusieurs Etats de l'Union Indienne, dont l'importance numérique est donnée en pourcentage de la population indienne, en raison de la croissance démographique : cachemiri (0,5%), panjabi (3%, et la moitié de la population soit 75 millions au Pakistan), sindhi (0,33%, 17 millions au Pakistan), marathi (8%) goujarati (5,36%), bengali (8,3%), oriya (3,7%), assamais (2%), konkani (0,3%), népali (0,3%), ourdou (5,6%). Selon le mode de recensement, les chiffres peuvent du reste différer. D'autant plus que les locuteurs tendent à déclarer comme langue maternelle celle à laquelle ils s'identifient culturellement : les locuteurs du bhili seraient passés d'un peu plus d'un million en 1971 à plus de 5 en 1991, ceux de l'hindoustani, proposé par Gandhi comme langue nationale car la plus parlée, estimés à 11 000 en 71.

Particulièrement variables sont les chiffres relatifs au hindi, étiquette qui selon Srivastava (1994) rassemble 331 langues et dialectes, dont les parlers du Bihar (bhojpuri, maithili, magahi, sadri) sont plus proches du bengali que du hindi standard, de l'Himalaya (les 'pahari' ou « montagnards »), et parfois du Rajasthan (marwari, mewari, malvi). Le hindi standard, langue urbaine, récente, n'est le fait que d'un très faible pourcentage de la population dite hindiphone, mais on emploie généralement le terme dans des acceptions variées, pour désigner soit l'ensemble des parlers des Etats de l'Uttar Pradesh, du Madhya Pradesh, de l'Haryana, de l'Himachal Pradesh, du Bihar (du Panjab au Bengale), soit le hindi standard (HS), soit la tradition littéraire et culturelle qui va du Panjab au Bengale. Comme les autres langues néo indo-aryennes (NIA), le hindi a connu trois stades : 1) le stade ancien, remontant à Amir Kushrau et aux ballades épiques en dingal (dialecte littéraire du Rajasthan) ou en maithili au 13^e s, puis aux chants lyriques en *sant bāṣā*, langue composite des poètes mystiques itinérants, dont une partie a été plus tard colligée dans le livre saint des Sikhs (*Guru Granth Sahib*) ; 2) le stade moyen ou médiéval, dominé par la littérature dévotionnelle, au 15-16^e s, krishnaïte d'un côté, en braj (dialecte occidental), ramaïte de l'autre, en awadhi (dialecte oriental) ; 3) le stade moderne, influencé par la colonisation britannique (19^e s.), voit naître la prose et la fiction. C'est à ce stade que se différencient hindi et ourdou, à partir de l'hindoustani commun, alors que les mêmes textes (Kushrau, Kabir) sont aussi considérés comme aux origines de la tradition littéraire ourdoue. Ce qu'on appelle vieil hindi est donc du braj, de l'awadhi, du dingal, du vieux maithili, du « dehlavī », le hindi moderne représentant le développement et la standardisation d'une lingua franca (hindoustani) dérivée de ces parlers. Cachemiri (avec la poétesse Lalla), bengali (dans les ballades *carya*), goujarati et marathi (avec Jnanesvari), bien que les compilations ne soient pas antérieures au 16^e s., sont aussi littérairement attestés à des dates analogues ; singhalais, marathi, oriya sont en outre attestés par des inscriptions entre le 9^e et le 11^e. Mais le stade ancien n'est pas attesté en

¹ (litt. : celui qui sacrifie correctement, et par là est victorieux)

panjabi ni sindhi, le patronnage des cours, important dans le développement des langues littéraires, ayant été moins efficace que dans la plaine du Gange ou le Rajasthan. Le stade moderne de même n'est pas contemporain entre bengali ou hindi et kashmiri où il commence à peine. Le NIA est dans son ensemble postérieur aux raids turko-afghans du 11-12^e s. et la période moyenne est contemporaine de la domination moghole, qui, avec la culture musulmane, introduit le persan comme langue juridique et administrative.

Certaines anciennes grandes langues littéraires, sont aujourd'hui des 'dialectes', dont certains parlés par moins de 30 000 locuteurs comme le braj (9 dingalophones en 1961), alors que d'autres 'dialectes' sont plus parlés en Inde que la 'langue' sindhi : en 1994, 23 millions parlaient le bhojpuri, 11 millions le chattisgarhi, 22 millions le maithili et dix millions le magahi, près de 5 millions le marwari, 2 millions le kumaoni ainsi que le garhwali. La reconnaissance par l'Académie des Lettres comme langue littéraire (bhojpuri, maithili, dogri (2 M), marwari), mais surtout par l'article VIII de la Constitution comme langue officielle, est déterminante pour le développement, la standardisation et la diffusion (medias et éducation), et conditionne de fait la différence entre langue et dialectes : la réalité sur le terrain est celle d'un continuum linguistique, dont les variantes géographiquement proches sont interintelligibles, et qu'il est parfois difficile de couper (ainsi le khandeshi ou ahirani a pu être décompté comme dialecte du marathi, du goujarati, du rajasthani, et le halbi, du marathi puis de l'oriya). Le siraiki au Pakistan a depuis peu un statut de langue distincte du panjabi (ainsi que du pothohari et du hindko plus au nord), comme du sindhi. A ces distinctions s'ajoute celle des communautés de locuteurs, aucune des langues IA dites tribales n'ayant pas (encore) de statut officiel (bien que le bhili soit parlé, entre Rajasthan, Maharashtra et Madhya Pradesh, par trois fois plus de locuteurs que le sindhi).

Outre les langues régionalement différenciées sur ce continuum, divers parlers se présentent discontinument : l'ourdou est une des deux langues officielles du Cachemire, du Bihar, mais est parlé aussi dans le sud de l'Inde sous le nom de dakhini (depuis 1347 avec les sultanats du Deccan), comme le lambani ou banjari, le shaurastri, deux parlers originaires du sud-ouest de la zone IA, témoins de l'importance et de l'ancienneté des migrations dans le Sous-continent, témoins privilégiés aussi des processus de contact et d'hybridation linguistique, qui ont par ailleurs affecté l'ensemble des langues indiennes (Montaut 1997 : 11-30). Le singhalais, issu du déplacement de populations bouddhistes de l'Inde du nord, aujourd'hui en contact avec le tamoul au Sri Lanka, en est un autre exemple. Le rromani ou tzigane représente un cas particulier, ses locuteurs, non sédentarisés, étant entrés en contact avec des ethnies et des langues très diverses (slaves, balkaniques, européennes). Diverses lingua franca se sont constituées pour surmonter les obstacles d'une situation multilingue comme le nagamais (structure assamaise et lexique naga), interlangue dans les Etats du Nord-Est à majorité linguistique tibéto-birmane, ou le hindi de bazar à Calcutta ou à Bombay, le sadri au Bihar et en Orissa, où pullulent les langues austro-asiatiques. Il faut se représenter que dans tous les états du centre de l'Inde, des langues austro-asiatiques mais aussi dravidiennes sont parlées avec les langues IA, qu'aucun Etat indien n'est monolingue, et moins d'un tiers des quelque 350 districts le sont, si l'on exclut les variantes dialectales. Les koinès observables dans les diaspora répondent différemment à une situation analogue mais dans un milieu dominé par un multilinguisme différent (français et créole à Maurice, fijien à Fiji, avec les parlers du Bihar et notamment le bhojpuri en lien avec l'origine des 'engagés' indiens recrutés dans les plantations au 19^e s.). Elles se distinguent du HS plus encore que les parlers régionaux d'origine qui leur ont servi de base.

L'un des problèmes posés par la standardisation des langues 'reconnues' et leur développement institutionnel est l'aggravation de la diglossie, (en bengali la variété haute, *sādhu bhāṣā*, s'appose à la *colit bhāṣā*, informelle) ; dans le cas du hindi, très diversifié régionalement, cette diglossie recoupe un bidialectalisme qui est un réel bilinguisme, les

parlers ‘bihari’ par exemple, étant linguistiquement plus distincts du HS que le panjabi ou l’ourdou. Elle est partiellement responsable du faible taux d’alphabétisation au Bihar et au Rajasthan (alors que Goa, beaucoup plus homogène linguistiquement, et le Maharashtra connaissent la plus forte alphabétisation en IA). Si le développement d’une langue standard, et surtout administrative et technique, très éloignée des dialectes, a affaibli la portée de la planification linguistique en matière d’alphabétisation, le hindi commun ne s’en est pas moins fortement diffusé en Inde, en partie grâce au cinéma commercial.

Les nombreuses classifications proposées depuis Grierson diffèrent dans l’interprétation des filiations génétiques et des affinités structurelles : le cachemiri, qu’il séparait de l’IA avec les langues dardes et kafires, est à présent considéré comme une de ses branches, la branche darde, comprenant aussi khowar, kohistani, shina, kalasha, pashai, contrairement aux langues kafires, mais une branche isolée, dont l’évolution a divergé de la majorité des autres langues NIA. Le romani, que Grierson groupait avec les langues dardes, se classe depuis Turner dans le groupe nord-ouest ou central de l’IA –des influences dardes s’y **retrouvent**, mais aussi sur les ‘pahari’ et les parlers rajasthanais. Les termes de ‘bihari’, ‘rajasthani’, ‘pahari’ ou ‘lahnda’, dus aussi à Grierson, recouvrant toute une variété de langues et dialectes, ont contribué à étayer la théorie des cercles concentriques correspondant à des invasions IA successives : la zone centrale, avec comme noyau le hindi occidental, l’ourdou, le panjabi, correspondrait au socle le plus ancien, la couronne extérieure, du marathi au bengali, à une seconde invasion. On admet aujourd’hui qu’il n’y a pas de lien entre les caractères linguistiques et l’arrivée supposée des groupes qui les parlent aujourd’hui, majoritairement hindous, mais aussi musulmans, sikhs, bouddhistes, jains, chrétiens, animistes (de nombreuses tribus locutrices de langues austro-asiatiques ont adopté des langues IA). Le tableau suivant situe les grandes langues et dialectes IA du sous-continent, d’après la synthèse des classifications détaillées dans Masica (1992 : 446-463) : les langues officielles sont soulignées, les langues nationales soulignées et en italique. On voit que le hindi standard comme l’ourdou (structurellement remarquablement analogue) est du côté des langues IA de l’ouest, supposées dérivées de prakrits occidentaux (sauraseni en particulier), alors que les parlers orientaux groupés sous le terme hindi (dont le bihari) sont du côté des langues IA de l’est, néomagadhéennes. En réalité, il y a une rupture assez radicale entre les divers parlers moyen indiens, tels qu’ils sont attestés par les prakrits puis les apabhramsha littéraires, et les langues néo indo-aryennes. Bien que nombre des parlers modernes soient régulièrement traités comme des dialectes (d’une grande langue régionale), les raisons en sont rarement linguistiques, et les dits dialectes du hindi comme l’awadhi ou le bhojpuri sont ici décrits avec les autres langues et non à la rubrique hindi.

kashmiri K	‘pahari’ occid dogrigarhwali	‘pahari’ oriental kumaoni <i>népali</i> N			<u>assamais</u> A
‘lahnda’ L siraiki	<u>panjabi</u> P	<i>hindi</i> occid <i>ourdou</i> O	hindi oriental awadhi Aw	‘bihari’ : maithili magahi puri	<u>bengali</u> B
<u>sindhi</u> S	‘rajasthani’ R marwari bhili khandeshi	braj HSbhoj		
	<u>goujarati</u> G	<u>marathi</u> M	chattisgarhi		<u>oriya</u> Or
		<u>konkani</u> Ko			
		dakhini			
				<i>singhalais</i> Sg	

2. écriture

Les écritures des langues indo-aryennes dérivent toutes de la même source, la brahmī (telle qu'on la trouve attestée dès le 3^e s., avec la *karosṭhī*, dans les inscriptions sur rocher et sur pilier d'Asoka). Une branche de la brahmī a abouti à l'écriture *Gupta* (4^e-5^e siècles), dont la version cursive, la *kuṭilā* ou « penchée » est à l'origine d'une part de la *devanāgarī* (« divine urbaine / métropolitaine ») et de ses dérivés, la *nāgarī* moderne, la *kaythī* (écriture de la caste des Kayastha, écrivains et clercs), la *bodhī* ou *bodiā* ou encore *saraf* en usage chez les commerçants du sud-ouest de l'Inde du nord, issue de la *kāythī*, d'autre part du 'proto-bengali', à l'origine de l'écriture maithili et bengali moderne (7^e s.) et de celles de l'assamais et de l'oriya (19^e s.). L'écriture du goujarati s'est développée au 19^e s. à partir de la *nāgarī* moderne, remplaçant la *modī*, écriture attribuée à un ministre de Shivaji (17^e s.) utilisée dans la région marathe et le Surat. Du rameau nord de la *Gupta* est née l'écriture *sharada* (ou *landā*, « du nord-ouest »), longtemps utilisée par les Brahmanes du Cachemire et dont dérive la *gurmukhī*, écriture du panjabi moderne attribuée au second guru sikh Angad (16^e s.). La *nāgarī* est aujourd'hui l'écriture officielle du hindi et de ses dialectes, des parlers rajasthani, du marathi, du népalī. Outre les divers dérivés de la *brahmī*, des écritures étrangères ont été utilisées (un alphabet de type khmer pour le chakma, dialecte bengali de Chittagong) ou le sont encore : le plus important de ces emprunts graphiques est l'écriture arabe, dite arabo-persane en Asie du Sud, dans sa variante *nasta'liq* (signes additionnels notant la rétroflexion), employé pour écrire l'ourdou, le sindhi, le kashmiri ; l'alphabet *tana*, utilisé pour écrire le dhivehi, en est aussi dérivé ; l'alphabet latin a d'emblée été choisi pour écrire le konkani, reconnu comme langue officielle en 1994, car il le distinguait du marathi, **encore qu'on** cherche à présent à le remplacer par un alphabet indigène ; inversement, des langues non indo-aryennes comme le santali, longtemps restées sans écriture, puis ayant adopté la *nāgarī*, cherchent à se doter d'un alphabet indigène (*olchitti*, *chitki*). Les écritures sont en effet de très forts symboles d'identité non seulement linguistique mais religieuse et régionale, et aujourd'hui dans la zone où se parlent hindi panjabi et ourdou, on tend à identifier écriture *nāgarī*-hindi-hindou, arabo-persan-ourdou-musulman, gurmukhi-panjabi-sikh, alors que jusqu'au début, voire au milieu du 20^e s., l'écriture arabo-persane était largement utilisée pour écrire le hindi et le panjabi : Inshā Allāh Khān a utilisé les caractères arabo-persans pour le texte de fiction réputé fondateur du hindi moderne (*Rānī ketkī kī kahānī*, 1803). Quant à la combinatoire de graphes, et notamment l'usage des diacritiques pour noter les voyelles après consonne, elle est dans tous les dérivés de la *brahmī* analogue à celle de la *nāgarī* à quelques détails près (les diacritiques circonscrits du bengali et de l'oriya pour certaines voyelles, comme en dravidien).

3. développements phonologiques

Les plus marquants, outre la complète phonologisation de la rétroflexion, sans doute polarisée par le substrat dravidien, se situent dans le prolongement du MIA : chute des voyelles finales et élision du a bref (schva), réduction des groupes consonantiques et amuïssement des consonnes intervocaliques. Quant à l'opposition des 3 sifflantes, déjà neutralisée en MIA en faveur de *s* à l'ouest et de *ś* à l'est (H *sun-* entendre, B *śun-*), elle se maintient dans les langues dardes (K).

La chute des consonnes finales en MIA a entraîné une dégradation du paradigme casuel qui aboutit en NIA à une modification fondamentale du système, un paradigme de postpositions ou de suffixes remplaçant l'ancien système flexionnel. Les voyelles brèves demeurées en finale disparaissent à leur tour (*rāga* > *rāg*, *Rāma* > *rām*)², sauf en maithili ou bhojpuri, qui conservent une voyelle finale ultrabrève, et en sindhi, où elle est distinctive

² ainsi parfois que les longues (jībhā > jībh langue), rātrī > rāttī > rāt nuit

(porteuse du genre notamment). Les diphtongues du sanscrit, sont réduites dès le MIA, produisant les voyelles longues e et o, abrégées devant CC, puis les voyelles ouvertes du HS □, □, mais celles-ci correspondant dans les dialectes et de nombreuses langues NIA à une diphtongaison (maithili), voire triphthongaison (magahi), parfois renforcée par un glide dans les variétés basses. Le glide est aussi fréquent entre deux V dans le hindi rural, alors que la séquence VV est admise en HS, sauf devant le -ā du p.p. (*bhāī* frère < *bhrat*², *kaī* plusieurs, mais *ā.y.ā* venu, *so.y.ā* endormi) ; les parlers non standards ont généralement VglideV (*āve*, *āye* pour *āe*, *kvī* ou *koy* pour *koī*).

Une innovation particulièrement frappante est l'harmonie vocalique en bengali, où une voyelle s'élève ou s'abaisse en fonction de la voyelle du morphème suffixé (*kena* acheter, *kini* j'achète), phénomène qu'on observe en télougou. Le cachemiri présente aussi une harmonie vocalique qu'on attribue parfois au même conditionnement morphologique (*bod*⁴ grand, *būd*ⁱⁱ grande, *bad*ⁱ grands, *baje* grandes, la voyelle finale surbrève, même non réalisée, conditionnant la nature de voyelle radicale).

Les consonnes intervocaliques se sont affaiblies puis amuïes, alors que les semi-consonnes initiales ont tendance à s'articuler comme la consonne correspondante : *catur* > *cār* quatre, *apara* autre > *avara* > *aur* et, *vīśati* > *bīs* vingt, *gopāla* > *gwālā* bouvier, *vivāh* > *byāh* mariage, *yah* > *jo* qui.

Les groupes consonantiques du sanscrit se sont simplifiés en moyen indien (C1C2 > C2C2 > C2), la voyelle précédante s'abrégeant au second stade si elle était longue à l'origine **VC1C2** > **VC1C2**³, puis subissant un allongement compensatoire lors de la simplification des géminées (*sapta* > *sāt* 7, *pārśva* > *passa* > *pās* à côté. Le panjabi et le sindhi font doublement exception en conservant des voyelles longues devant les géminées (*āttā* farine, *gājjar* carotte) et en échappant à l'allongement compensatoire (*agg*, « feu » < *agni*, *satt* « sept » < *sapta*, *pillā* « jaune », qui sont respectivement *āg*, *sāt*, *pīlā* en hindi). La combinaison kṣ ne s'est préservée qu'au nord-ouest (K), a donné kk(h) ailleurs, ou chh, **voire** ts, ps.

Le hindi semble faire exception par l'abondance des groupes autorisés, alors qu'on ne devrait trouver, dans une évolution régulière indigène, que C+semi-voyelle à l'initiale ; ils s'expliquent par des emprunts (au sanscrit strī « femme », à l'arabe *qism* sorte, à d'autres langues, *lift*). L'emprunt explique aussi le maintien de l'opposition *s/ś*, de la séquence *ri* < ², : *kṣamā* pardon, *m²tyu* mort, *śanti* paix, *putr* fils. La néologie moderne contribue à préserver ces groupes (*rāṣṭrabhāṣā* langue nationale), mais les variantes régionales non standardisées ne les présentent généralement pas (*kānh* pour K²ṣṇa, *būḥā* « vieux » pour *vṛddha* (> *buḍha*), *cānd* ou *candar* pour *candr* lune, *sādī* et non *sādī*, mariage, *bhāsā* pour *bhāṣā* langue, *saram* et non *śarm*, honte), et le parler non soutenu même en HS restaure des voyelles épenthétiques (*dharam* pour *dharm*, *śaram*, pour *śarm*, *qisam* pour *qism*, *iskūl* pour *skūl*).

Les dialectes représentent donc un conservatoire des phases diachroniques dépassées, ou renversées par l'emprunt au sanscrit, dans le standard, lequel est phonétiquement à la fois archaïque du fait de cet emprunt massif et novateur du fait du statut nouveau de la voyelle a.

Parmi les évolutions spéciales, celles du *ṇ* (spontanément cérébralisé en MIA), qui s'est maintenu dans la plupart des parlers plus occidentaux que le hindi ('R', G, M, P, S, garhwali et même haryanvi ont *pāṇī* eau) et a rejoint la dentale à l'est (HS, 'bihari', B, A ont *pānī*), l'oriya étant à part (*ṇ*). La distribution de *ḷ* par rapport à *l* est à peu près analogue (*phaḷ*, fruit, à l'ouest, *phal* à l'est) mais elle se complique de variantes en -r (*phar* en S, maithili, bhojpuri). Le kashmiri, qui a *n* et *l* contrairement aux parlers occidentaux, a par ailleurs développé une série de consonnes dites palatalisées (originellement consonne + i/e).

La particularité apparemment la plus atypique en IA est l'apparition d'un système de tons en panjabi : les aspirées voisées se sont assourdies et déaspirées à l'initiale et ont donné

³ V(longue ou brève)C1C2 > V(brève)C1C2

un ton bas sur la voyelle suivante, elles se sont déaspirées en finale et ont donné un ton montant sur la voyelle précédente (à H *ghoṛā* cheval, correspond P *koṛā*, à *bhāi* H *pāi*, à *dūdh* H lait *dūd*, à *jībh* H langue *jīb*) Noter les tons. La dé-aspiration des voisées en kashmiri n'engendre pas de tons, ni en singhalais où elle est liée au contact avec le dravidien (Sk *sukha* > *suva* bien-être, *pote* « livre » vs. *pothā hindi*), ou en dakhini (*bī* aussi, vs HS *bhī*).

4. morphologie

L'érosion phonétique des finales est responsable du passage d'un système hautement flexionnel (sanskrit) à un système de postpositions ou de suffixes. La frontière entre morphologie et syntaxe en NIA est d'autant plus délicate et moins pertinente que c'est en vertu de conventions purement graphiques que les marques de cas sont suffixées en marathi et détachées en hindi (mais non dans les parlers non standard récemment écrits), ces marques étant souvent d'origine lexicale (adverbe, verbe, nom), plus rarement casuelle, et qu'elles se suffixent sur une forme du mot parfois distincte (d'origine flexionnelle). De même les auxiliaires verbaux sont séparés en HS mais non en dakhini (*atū* vs HS *ātā hū* je vais), ils sont amalgamés dans certains tiroirs verbaux du bengali : *colchi* et *colechilam* correspondent à HS *cal rahā hū* je marche, et *calā thā j* j'avais marché. Le futur 'simple' du HS était jusqu'au dix-neuvième siècle perçu comme analytique, permettant l'insertion d'une particule entre les morphèmes de personne et de temps (*kar.ū.hī.g.ā* faire-1s-juste-futur-ms, je ferai), mais *karūgā* aujourd'hui ne la permet plus. Traiter la morphologie sur la base du mot graphique ne rendrait compte uniment ni de la structure morphologique des langues IA dans leur ensemble, ni du hindi même. Il faut donc considérer ensemble les marques casuelles (flexions, suffixes et postpositions), comme les marques aspecto-temporelles (flexions, suffixes, mais aussi auxiliaires, verbes supports), dans la mesure où certaines langues les intègrent pleinement à la morphologie, d'autres imparfaitement.

D'une manière générale, la flexion nominale est limitée en NIA à deux formes, indépendante et dépendante (devant postposition ou suffixe). Les 'degrés' de la base dans les 'dialectes' orientaux du hindi (*ghoṛa*, *ghoṛā* *ghoṛwā* *ghoṛawā* « cheval », en Bh, maithili, magahi) sont d'une autre nature et constituent une variation pragmatique (+/- familier, respectueux) ou « expressive », qui s'obtient dans les parlers occidentaux (R) par des suffixes (-k-, -l-, -ṛ-, mais aussi -k- en Aw).

Une des désinences casuelles flexionnelles (aptées à marquer la fonction en l'absence de postposition) les mieux préservées en dehors du HS est celle du locatif en *-i* (vieil hindi, marathi moderne) ; *-(t)e* bengali relève plutôt de la suffixation. Le cachemiri, le marathi ont conservé davantage de traits flexionnels que le HS, le HS davantage que sa variante méridionale (ils sont affaiblis en dakhini, disparus dans le hindi de bazar et les koinés des diasporas) et que les langues du groupe oriental.

Les génitifs suffixaux du bengali (*-r*) ont en fait la même origine (*kera* participe de la base verbale *kar*) que le *kā* servant de génitif en hindi, et sont caractéristiques des parlers orientaux (chattisgarhi, maithili, bhojpuri), mais aussi du rromani (*-kero/kiro*). Les formes possessives du HS pour la 1^e et la 2^e pers (*merā*, *terā*, *hamārā*, *tumhārā*) sont de même origine.

En ourdou formel, un substitut du génitif dans la séquence N2N1 est le procédé persan de la jonction par ézafé (N1-e-N2), renversant ainsi l'ordre habituel en IA N2-gen-N1. Le même renversement s'observe en rromani, pour des raisons différentes (contact avec les langues slaves, balkaniques, germaniques) avec l'émergence d'un sous système de

prépositions (*prdal* pour, *ki* à), concurrençant les suffixes casuels : *Praha-θe* ou *ki Praha*, à Prague⁴.

Quant aux marques de genre et de nombre, réduites dès le MIA tardif (perte du duel), elles font état aujourd'hui d'un continuum de la grammaticalisation du genre qui décroît de l'ouest à l'est : 3 genres en M, G, S, 2 en H/Ou, P, R, genre non marqué à l'est (B, A, Or, Sg). Ce continuum recoupe partiellement celui de la déclinaison et du nombre, et confère un caractère clairement agglutinant aux parlers orientaux. Le nombre y est marqué soit par des suffixes soit par des termes collectifs distincts pour les animés et les inanimés (B *-erā gula/i*, chattisgarhi *man, sab*, maithili *lokan, sab*, O *man, sob*) et ne subsiste pas dans le paradigme verbal. Par ailleurs la catégorie morphologique de la définitude, ignorée en sanscrit et MIA, est pertinente à l'est (bengali *-ṭā l-ṭī, -khāṇā l-khānī*, présents dans les dialectes orientaux du hindi et parfois considérés comme classificateurs du fait de leur cooccurrence avec les numéraux ; 'article' défini du singhalais *-ek*), alors que les langues occidentales marquent la définitude de façon syntaxique (datif de l'objet) ou discursive, et ont des pluriels partiellement flexionnels, y compris par umlaut (kumaoni : *celo, cyālā*, fils, *ghoṛo/ghwārā* cheval/aux, konkani : *por / paura* garçon/s).

L'appauvrissement de la flexion nominale en hindi méridional est accentué par l'environnement dravidien, poussant vers un type agglutinant (avec fragilisation du genre et de la flexion). Un pronom de 1^{ère} personne pl. inclusif s'y est développé ainsi qu'en M, G, et dans le bombay-hindi (*apan*), sous l'influence des langues dravidiennes, qui, à la différence de l'IA, distinguent incl. et excl.

Dans le syntagme verbal, deux formations dessinent des isoglosses significatives. La première est celle du futur avec quatre expressions essentielles : l'ancien futur sigmatique du sanscrit s'est conservé à l'ouest et au centre (-s-, -h-, -f- : G, R, S, Aw, Bh), le futur en -g-, <*gach* aller, s'est développé au centre (H P Ou), et les formations en -l- (< *lag*, toucher) au nord (parlers himalayens dont N) et au sud (M, Ko), les formations en -b- (< participe d'obligation *-tavya*) dominant à l'est (B, Aw, Bh, maithili, Or, A). La seconde isoglosse est celle de l'accompli, issu du p.p. sanscrit *-ita* (-ṭ- en HS/Ou), avec la variante -y- systématique à l'ouest (sindhi, K, P), alors que le morphème -l-, ancien suffixe diminutif, caractérise une vaste isoglosse allant du groupe oriental (Bh, maithili, B, A, Or) au sud-ouest (M, Ko), le goujarati ayant les deux formations. Le p.a. issu du p.a. sanscrit *-anta*, s'est moins différencié, ayant des formes en -nd-, -d- au nord-ouest (P, 'pahari'), en -t- en HS, en *-at, -it* en braj et awadhi. Le participe dit conjonctif a gardé sa forme ancienne -i <scr *-ya*) dans de nombreux dialectes hindi, mais il est en *-ūn* en marathi. Le verbe être a diverses formes, liées aux bases sanscrites à leur origine (BHŪ>*ho*, STHĀ>*thā, thyā*) réparties selon les temps, mais la base en *ch* (*ach*)- est aussi fréquente, alors que *bāṭ* est spécifique du bhojpuri. La 'double' dérivation causative (base/base+a ou umlaut/base +va) permettant de convertir une base intransitive en transitif puis causatif ou factitif est une structure pan-indienne, attestée en tamoul ancien mais non en sanscrit. Le nombre des locutions verbo-nominales par rapport aux verbes simples est aussi un phénomène panindien.

Une innovation générale du NIA comme des autres langues indiennes modernes est l'emploi de verbes de mouvement semi-auxiliarisés dans des fonctions de perfectif, d'expressivité, et d'orientation de la diathèse. Leur grammaticalisation varie selon les langues. Les parlers romani en contact avec les langues slaves ont intégré des préfixes perfectivants (*po-*) dans cette fonction.

⁴ De même Moskvaθar, de Moscou est concurrencé par andar-i Moskva, où l'adverbe IA se combine avec l'ezafé pour renverser la séquence

5. Syntaxe

L'énoncé élémentaire dans les langues NIA a acquis trois particularités essentielles et panindiennes : un ordre non marqué SOV (le kashmiri et le rromani faisant exception) ; l'expression oblique du 'sujet' d'un prédicat non actif (généralement X-dat peur/savoir/sentir/avoir, X-gen en B, Xdat/gen/loc en HPOu), « avoir » n'ayant pas d'équivalent lexical en IA ; un marquage différentiel (généralement par le marqueur de datif) des objets humains ou inanimés spécifique (acc. marqué, que le dravidien associe à un marqueur spécifique). Dans les langues de l'ouest et du nord (H, P, Ou, K, G, M, 'R', 'Pahari', N) l'énoncé transitif est en outre ergatif à l'accompli.

Ce développement historique majeur prolonge l'usage prédicatif du p.p. (-*ita*) en sanscrit avec l'agent à l'instrumental (*mayā tat k²tam* /1sg-instr/ ceci/ fait/ je l'ai fait). Ce renouvellement périphrastique du parfait évoque celui du latin par la périphrase « possessive » *mihi id factum*, une prédication de localisation, avant qu'il ne se transforme en structure active avec le verbe avoir. Alors que les langues occidentales renforçaient le tour par un marqueur spécifique (*ne/ni* H,M , P, O, R, *la* 'pahari', N), les langues orientales (B, Or, A) comme les parlers du Bihar retrouvaient une structure nom-acc. Aujourd'hui les formes de l'énoncé ergatif varient grandement (agent oblique sans marqueur spécifique, accord avec l'objet marqué en 'rajasthani' occidental mais non en HPOu, accord avec le sujet et le patient en M à la seconde personne). La variante méridionale du H (Da) s'est, elle, dé-ergativée par contact avec le dravidien, ainsi probablement que le singhalais si on admet que le pali représente un état pré-ergatif. Quant au kashmiri, ergatif, il présente un système complexe de clitiques qui indexent sur le verbe les participants principaux de l'énoncé, comme le siraiki (et les langues munda, génétiquement non apparentées). Maithili et magahi ont aussi un accord avec plusieurs participants, mais qui doit plus au substrat munda qu'à l'ergativité, les verbes intransitifs et inaccomplis pouvant indexer plusieurs participants animés.

La phrase complexe préfère les propositions dépendantes participiales ou nominales, mais a bien maintenu l'ancien système corrélatif du sanscrit : la base relative *j-* (<scr *y-*) introduit le premier membre du diptyque auquel est corrélé le second, introduit par un déictique (*v/-t-*). Sur le modèle de la relative *jo...vah*, lequel... celui-ci (= qui), on a *jab... tab*, quand... alors, *jahã... vahã* où... là, *yadi...to*, si... alors, etc. Ce système est si prégnant que la complétive (empruntée au persan : *ki* que) est souvent anaphorisée ou cataphorisée par un déictique.

Par contact la dakhini a calqué un emploi spécifiquement dravidien du « quotatif » (forme figée du verbe 'dire') qui, postposé à une proposition dépendante, et devant un verbe principal, fournit toutes sortes de relations d'hypotaxe (complétive, causale, consécutive, conditionnelle), en conformité avec l'ordre strict élément régi-élément recteur qui prévaut en dravidien. Ainsi en marathi (*mhaṇūn*), bengali (*bole*), oriya (*buli*) singhalais et népalais. Cet ordre est corollaire de la postposition de la particule interrogative qui domine la phrase, final en dravidien et dans les langues IA à quotatif, initial en HPOu. De même, alors que HPOu ont un réfléchi lié dans sa proposition, en cela typiquement indo-aryen, le marathi a aussi, à côté du réfléchi strict (*swatāh*), un réfléchi à longue distance (*apliya*) qui coréfère avec le sujet d'une proposition supérieure, fait typiquement dravidien. L'existence d'une forme verbale à valeur médiative (« évidentielle ») en kalasha, khowar, mais aussi en népalais, peut évoquer une influence sino-tibétaine. Le bengali, l'oriya, le singhalais, expriment la phrase copulative sous forme nominale, et ont une forme spéciale de copule négative, la négation du verbe existentiel étant distincte, comme en dravidien. Le contact a donc exercé un rôle aussi déterminant dans le changement syntaxique que morphologique et lexical.

7. lexique

On a longtemps distingué les mots *tatsam*, « tels qu'ils sont » (en sanscrit), ou du moins tels qu'ils sont écrits (scr *abhyāsa* exercice se prononçant *abhyās* en H, □*bhaeś* en B, *abhax* en A, scr *jñāna* connaissance, *gyān* en HS, *ḍnyān* en marathi, *gaen* en bengali), des mots *tadbhav* (tels qu'ils sont devenus) comme *sāt* (< *sapta*), ou *jān* (H) et *dzān* (M), < *jñāna*, et des mots *deśī*, indigènes. Les deux dernières catégories ne diffèrent en réalité que par le degré de lisibilité de l'étymon, les mots *deśī* étant souvent des *tadbhav* plus obscurs. Quant aux *tatsam*, le fait qu'on les trouve en sanscrit ne signifie pas nécessairement qu'ils soient indo-aryens : de très nombreux substrats et adstrats ont enrichi le vocabulaire védique : d'abord de l'Asie Centrale, Bactriane et Margiane, comme le nom du chameau (*uṣṭra*, > *ūṣṭ* en NIA), de la marque (*ling*), puis para mounda (préfixés : *ku-māra* « jeune homme », *pu-rusha* « homme ») dès les livres anciens du Veda, ensuite dravidien dans les livres plus tardifs (*ukha* hanche). Ces termes non indo-aryens (*phal* « fruit » : dravidien), dont l'origine n'est pas toujours encore claire (*punya* vertueux : austro-asiatique ?), empruntés très anciennement, constituent une couche non négligeable du vocabulaire, notamment agricole. Les emprunts ultérieurs doivent peu aux premiers conquérants turco-afghans du 10^e au 11^e s. (türk : *begum*) et aux colons portugais (chou-fleur : H. *ghobī*, Sg *gova*, M *kobī*, *almārī* armoire, *kamarā* pièce), beaucoup aux moghols (persan : *pyāz* oignon, *zamīn* terre, arabe : *urdū* camp, armée, *qānūn* loi, *zīkr* mention, *lekin* mais, *aurat* femme) entre le 15 et le 18^e s. puis aux britanniques (anglais *lift*, *trak*, *kār*, *sāin*, *sāikil*).

L'élément sanscrit (*tatsam*) représente une particularité marquante du NIA par rapport au moyen indien. Au 19^e s. **Beames** en dénombreait une plus forte proportion dans les langues orientales qu'occidentales mais la sanscritisation du hindi standard a renversé les choses. Inversement c'est l'arabisation qui distingue l'ourdou moderne après la persanisation (ourdou littéraire classique). Beaucoup de verbes étrangers (persans comme anglais) ont été morpho-syntaxiquement indianisés par la structure des locutions verbo-nominales, structure pan-indienne très productive dans l'enrichissement du lexique : comme on a *bāt karnā* (< scr *varta / v²ta*), on a *band karna* fermer (Pe) *xatm karnā* finir (Ar), *join karnā*, *fon karnā*, et même *on karnā* mettre en marche (angl).

Egalement panindienne est l'importance dans le stock lexical des formations à base onomatopéique ou « expressive » (*jhaṭak* secousse), souvent redoublées (B,M : *tharthar* trembler H, P : *jhamjham* briller, P, N, B, A : *dagmag* trembler, hésiter), et les mots dits à écho, qui élargissent le champ notionnel (H *khānā-vānā* nourriture etc., P *rotī-śotī* pain etc.), deux procédés toujours très productifs. Un autre trait pan-indien est l'exploitation systématique de l'opposition transitif/intransitif sur la même base prédicative (soit par dérivation causative, soit, dans les locutions verbo-nominales, par le choix du verbe support statif ou actif). Enfin, toutes les langues NIA ont en commun un grand nombre de verbes simples (manger *khā*, boire *pī*, entendre *sun*, etc.). Les faits lexicaux corroborent donc la convergence remarquable qui a conféré à l'IA moderne les traits structurels panindiens suivants, absents des langues européennes : rétroflexion, ordre des termes, marquage de l'objet animé ou spécifique, formation des causatifs, verbes sériels perfectivants, importance des locutions verbo-nominales, datif de l'expérient.

Beames J., [1872-9] 1970, *A Comparative Grammar of the Modern Aryan Languages of India*, New-Delhi, Munshiram Manoharlal.

Bloch Jules, 1934, *L'Indo-aryen du Veda aux temps modernes*, Paris, Maisonneuve

Cardona G. & Dh. Jain (eds.), 2003, *The Indo-Aryan Languages*, London-New-York, Routledge.

Deshpande Madhav. & Hook Peter E. (eds.), 1979, *Aryan and non aryan in India*, Ann Arbor, University of Michigan Press

Grierson George A., 1903-28, *Linguistic Survey of India* (11v. sur 19), Calcutta
Masica Colin, 1992, *Indo-Aryan languages*, Cambridge University press
Montaut Annie (ed.), 1997, *Les langues d'Asie du Sud*, Paris, Ophrys (Faits de Langue 10)
Montaut A., 2004, *Hindi Grammar*, Munchen, Lincom Europa (dialectes traités)
Srivastava R.N., 1994, *Bi/multilingualism*, New-Delhi, Kalinga.