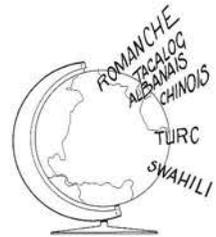


# Glossaire des langues



## Présentation

### Comment le Glossaire des langues est-il constitué ?

Le *Glossaire des langues* présente, par ordre alphabétique, toutes les langues qui sont utilisées – à un titre ou à un autre – dans les activités des deux volumes EOLE.

Chaque langue fait l'objet d'un court descriptif indiquant :

- son appartenance (ou non) à une *famille de langues* et sa parenté avec d'autres langues,
- sa localisation géographique (seuls les principaux pays sont cités),
- le nombre de locuteurs la parlant en tant que langue première (langue maternelle),
- son système d'écriture.

Quelques caractéristiques spécifiques de la langue sont également présentées : par exemple un aspect de son évolution historique, des questions socio-linguistiques en relation avec son statut ou des usages particuliers qui lui sont liés, une particularité de son fonctionnement grammatical ou de son système phonologique, etc.

### Glossaire des langues, Lexique plurilingue et prononciation

Des indications plus précises à propos de la lecture et de la prononciation des différentes langues ainsi que des principes qui ont guidé la transcription des mots figurant dans les activités et/ou dans le *Lexique plurilingue* sont fournies dans l'annexe *Quelques indications pour la lecture et la prononciation des langues*.

Les 20 langues que l'on retrouve dans le *Lexique plurilingue* sont signalées ici par un astérisque.

### Des informations à prendre avec prudence

Il est clair que le *Glossaire des langues* ne peut pas offrir une information exhaustive et définitive sur les langues présentées. En effet, certaines données sont difficilement accessibles, d'autres sont en évolution constante, d'autres encore varient fortement d'une source à une autre... Par exemple, le nombre de locuteurs d'une

langue peut se révéler extrêmement délicat à évaluer (absence de recensements récents, migrations de population, difficulté à savoir si, pour certaines populations, il s'agit d'une langue maternelle ou d'une langue seconde, etc.). La localisation géographique des langues pose également certains problèmes : en effet, les langues ne s'arrêtent pas aux frontières politiques des pays, elles les ignorent et les traversent sans cesse, entre autres lorsqu'elles accompagnent les migrants qui s'installent dans un autre pays.

Par ailleurs, nous n'ignorons pas que toute description d'une situation linguistique a des implications identitaires et politiques pour les lecteurs concernés, en particulier lorsque le statut de leur langue est problématique, voire objet de conflit. Excepté quelques remarques ponctuelles, nous avons cependant choisi de privilégier des informations moins « sensibles » qui, à notre avis, sont à la fois intéressantes – pour les enseignants et les élèves – et utilisables dans le cadre des activités.

Pour ces différentes raisons, nous invitons les lecteurs de ce Glossaire à prendre les informations y figurant avec toute la prudence nécessaire.

### Les sources principales du Glossaire

Les informations que vous trouverez dans le *Glossaire* sont en grande partie issues :

- de l'ouvrage de M. Malherbe, *Les langages de l'humanité. Une encyclopédie des 3000 langues parlées dans le monde*. (édité chez Robert Laffont, Paris, 1995),
- du livre *The Languages of the World*, de K. Katzner (édité chez Routledge, Londres, 1995),
- du livre *Langues sans frontières. A la découverte des langues de l'Europe*, de G. Kersaudy (Editions Autrement, Paris, 2001),
- du site internet de Jacques Leclerc, *Les grandes familles linguistiques du monde*, Québec, TLFQ, Université Laval ([www.tlfq.ulaval.ca/axl/](http://www.tlfq.ulaval.ca/axl/)),
- du site internet *Ethnologue: Languages of the world* ([www.ethnologue.com](http://www.ethnologue.com)) (en particulier pour le nombre de locuteurs par langue).

Lorsque d'autres sources sont utilisées, celles-ci sont mentionnées sous le texte descriptif de la langue concernée.



L'anglais connaît de ce fait d'importantes variations selon les pays où il est parlé (de manière quelque peu « ludique » et sans être systématiques, ces variations peuvent même parfois toucher le système graphique : *night* en anglais d'Angleterre, *nite* aux Etats-Unis...).

L'anglais est la plus latine des langues germaniques ! Près de deux tiers de son vocabulaire, en effet, est d'origine latine, la plupart de ces mots ayant transité par le français. De très nombreux emprunts sont par ailleurs revenus au français plus tardivement (ex. : *conter fleurette* → *flirt* → *flirter*).

Langue réputée simple d'un point de vue grammatical, l'anglais possède en revanche un vocabulaire très étendu et un nombre très élevé de locutions et autres formes idiomatiques pas toujours transparentes qu'il s'agit de mémoriser. Ses systèmes phonétique et graphophonétique sont également particulièrement complexes.

## \* Arabe

العربية

### Famille

langue sémitique

### Localisation

Algérie, Arabie Saoudite, Barhein, Yémen, Oman, Emirats Arabes Unis, Koweït, Qatar, Irak, Syrie, Jordanie, Liban, Egypte, Libye, Tunisie, Maroc, Mauritanie, Soudan, Israël, Sahara Occidental,...

### Nombre de locuteurs

environ 170 millions

### Système d'écriture

alphabet arabe (s'écrit de droite à gauche)

La langue arabe est l'une des langues les plus parlées au monde et, depuis 1974, elle est langue officielle aux Nations-Unies. L'arabe est devenu, grâce à l'Islam dont il est la langue liturgique, la langue sémitique la plus répandue (l'hébreu, l'araméen, l'amharique notamment font également partie de cette famille). En outre, en raison de cette influence culturelle et religieuse, les langues des peuples musulmans non arabophones, comme le persan, le turc ou le malais ont intégré de nombreux mots arabes.

Langue sémitique, l'arabe partage avec les autres langues du groupe un important stock de vocabulaire : par exemple, le mot *salam* en arabe (qui veut dire « paix ») est très proche du mot hébreu *shalom* ; les noms des chiffres montrent également une grande ressemblance dans ces deux langues (*arba'* – *arba* (4), *khamisa* – *khamesh* (5)...).

Les variétés de l'arabe sont nombreuses et il est parfois difficile pour des personnes arabophones de pays différents de se comprendre. Cependant, il existe une forme d'arabe moderne largement commune à l'ensemble des pays, excepté quelques variations de prononciation : l'arabe dit « moderne », construit à partir de l'arabe « littéral » ou « classique », qui est entre autres l'arabe de la presse et des médias. Il est prudent ainsi, lorsqu'on parle de l'arabe, d'ajouter un adjectif qui précise duquel il s'agit (moderne, littéral, libanais, marocain... !)

L'écriture arabe utilise l'alphabet arabe riche de 28 lettres et s'écrit de droite à gauche. Les lettres se lient à l'intérieur des mots et leur forme se modifie selon leur position dans le mot. Toutes les variétés de l'arabe ont au moins 3 voyelles longues /i/, /a/, /u/ et beaucoup possèdent les 3 brèves correspondantes. Le mythe souvent répandu d'une « langue sans voyelles » repose en fait sur une confusion de l'écrit et de l'oral. En effet, il existe deux variantes de l'écriture arabe : l'une dite « vocalisée » qui note soigneusement (par des signes diacritiques) les voyelles, longues ou brèves, et l'autre dite

« non vocalisée », qui ne note que les voyelles longues (et qui est plus courante dans la presse, les livres...). Dans ce cas, il faut inférer le sens du mot à partir de son organisation consonantique et du contexte : comme dans d'autres langues sémitiques, les mots sont le plus souvent formés à partir d'une racine de trois consonnes qu'on précise par différents procédés de dérivation : ainsi k.t.b, *écrire*, donne *mektub* (écrit), *katib* (écrivain), *maktab* (bureau), etc.

La grammaire de l'arabe se caractérise entre autres par l'existence – en plus du singulier et du pluriel – d'un troisième nombre, appelé *duel*, qui est utilisé pour renvoyer à deux unités.

## Arménien

հայերէն

### Famille

langue « isolée » de la famille indo-européenne

### Localisation

Arménie, Géorgie, Azerbaïdjan et diaspora (Iran, Syrie, Liban, Egypte, USA, Turquie, France, Irak, ...)

### Nombre de locuteurs

environ 5 millions

### Système d'écriture

alphabet arménien (s'écrit de gauche à droite)

L'arménien est langue officielle en Arménie ; mais il est également parlé par une importante diaspora – principalement issue des persécutions subies par les Arméniens de Turquie – qui cherche à en maintenir l'usage et à en préserver l'héritage culturel. Ces diverses communautés, disséminées dans de multiples pays, n'ont jamais cessé de publier de nombreux journaux et revues en arménien.

On peut distinguer aujourd'hui deux dialectes arméniens : celui parlé en Arménie (dit arménien oriental) et celui parlé par la diaspora (dit arménien occidental). Les Arméniens se désignent par le nom de Hay ; leur pays est l'Hayasdan, leur langue l'hayaren.

Bien que la langue arménienne ne puisse être valablement rattachée à aucune famille à l'intérieur du groupe indo-européen, certaines de ses caractéristiques la rapprochent du grec. Elle a également emprunté de très nombreux mots au *pashtou* et au *persan*.

L'alphabet arménien comprend 32 consonnes et 6 voyelles. Il a été élaboré au V<sup>e</sup> siècle par un moine nommé Mesrop Mashtots. Grammaticalement, l'arménien est une langue à 4 cas, qui contient des postpositions (et non des prépositions), qui place le verbe en fin de proposition et l'article après le nom.

## Aymara – aymara

### Famille

l'aymara constitue une famille à lui tout seul

### Localisation

Pérou, Bolivie

### Nombre de locuteurs

environ 2 millions (3<sup>e</sup> langue amérindienne en Amérique latine après le *quechua* et le *guarani*)

### Système d'écriture

alphabet latin

Le peuple aymara est l'un des plus importants d'Amérique latine. Son territoire se situe sur les hauts plateaux andins autour du lac

## \* Albanais – *shqip*

### Famille

langue « isolée » de la famille indo-européenne (ce n'est pas une langue slave)

### Localisation

Albanie, Macédoine, Kosovo, Monténégro et quelques petites communautés en Grèce et en Italie

### Nombre de locuteurs

environ 4 millions

### Système d'écriture

alphabet latin

La langue albanaise, qui s'est un moment écrite avec l'alphabet grec, utilise l'alphabet latin depuis 1908. Elle présente deux variétés linguistiques principales, parlées dans des proportions plus ou moins égales : le *guègue* et le *tosque*. Le guègue est parlé au nord de l'Albanie, ainsi qu'au Monténégro, au Kosovo et en Macédoine. Quant au tosqe, il est employé au sud de l'Albanie, mais aussi en Grèce et dans la région macédonienne du lac de Prespa. Ce n'est que récemment, après la dernière guerre, que l'orthographe de l'albanais a été normalisée, essentiellement à partir du tosqe, contribuant ainsi à la création d'une langue littéraire, laquelle semble avoir été acceptée par l'ensemble de la population. Les Albanais dénomment leur langue *shqip* et leur pays *Shqipëria*, termes dont la signification est liée à l'aigle de leurs armoiries.

D'origine indo-européenne, mais ayant évolué de manière spécifique, le vocabulaire albanais contient de nombreux emprunts au grec, au turc, aux langues slaves, au latin et à l'italien, mais aussi certains mots d'origine inconnue. L'albanais est une langue à cinq cas, comportant donc un système de déclinaison. Il possède également des déterminants articles et possessifs qui sont suffixés au nom.

## \* Allemand – *Deutsch*

### Famille

langue germanique (famille indo-européenne)

### Localisation

langue officielle en Allemagne, Autriche, Suisse, Luxembourg, Haut-Adige (Italie), Liechtenstein. Variétés dialectales également parlées dans l'est de la Belgique, en Alsace et en Lorraine.

### Nombre de locuteurs

environ 100 millions

### Système d'écriture

alphabet latin

L'allemand est la langue la plus parlée de la Communauté européenne. Sa diffusion et son prestige internationaux sont en revanche bien moindres que pour l'anglais, l'espagnol, le français et le portugais. L'allemand écrit (« Schriftdeutsch ») est unifié ; en revanche, il existe de nombreuses variétés dialectales orales, au point que l'intercompréhension entre elles peut parfois poser problème (cf. suisse-allemand).

L'allemand s'est longtemps écrit en écriture gothique, système de calligraphie utilisé depuis le XIII<sup>e</sup> siècle par les scribes et les moines de toute l'Europe, dont les caractères correspondent aux lettres latines, mais sont particulièrement travaillés (crochets, angles...). Après la Deuxième Guerre mondiale, l'alphabet latin l'a largement supplanté.

L'allemand conserve toutefois la particularité d'être la seule langue où tous les noms commencent par une majuscule. On relèvera également qu'il permet de former un nombre particulièrement élevé de mots composés formés par juxtaposition de mots réunis en une seule unité.

## Amharique



### Famille

langue sémitique (même famille que l'arabe, l'hébreu, le tigrinia, etc.)

### Localisation

Ethiopie (Addis Abeba et le nord de l'Ethiopie)

### Nombre de locuteurs

environ 18 millions

### Système d'écriture

alphabet éthiopien (écrit de gauche à droite)

L'amharique est la langue nationale de l'Ethiopie, parlée par un tiers de la population (on trouve, parmi les autres langues le tigrinia, le galla et le somali). D'autres langues sémitiques de la branche éthiopienne, comme le tigrinia et le tigréen, en sont proches parentes. Ces langues s'écrivent avec un alphabet spécifique, parent des anciennes écritures sud-arabiques (aujourd'hui disparues) et fondé sur le même principe qu'utilisent les alphabets indiens (signes complémentaires attachés aux consonnes pour marquer les voyelles).

De manière générale, les langues sémitiques (terme créé par les linguistes à partir de Sem, fils du patriarche Noé, frère de Cham et père d'Aram, dans le but d'exprimer les liens de parenté entre ces langues) partagent de nombreuses racines lexicales communes.

## \* Anglais – *english*

### Famille

langue germanique (famille indo-européenne)

### Localisation

– *Langue officielle et langue première pour une grande partie de la population* : Afrique du Sud, Australie, Canada, Etats-Unis, Irlande, Nouvelle Zélande, Royaume Uni.

– *Langue officielle, mais créoles de souche anglaise parlés par la population* : Jamaïque, Petites Antilles, Surinam, Guyana.

– *Langue officielle, mais langue seconde pour la majorité de la population* : Belize, Botswana, Cameroun, Fidji, Gambie, Ghana, Inde, Kenya, Kiribati, Lesotho, Liberia, Malawi, Maldives, Malte, Maurice, Nauru, Nigéria, Ouganda, Pakistan, Papouasie, Philippines, Salomon, Samoa, Seychelles, Sierra Leone, Singapour, Sri Lanka, Swaziland, Tonga, Tuvalu, Vanuatu, Zambie, Zimbabwe.

### Nombre de locuteurs

environ 320 millions

### Système d'écriture

alphabet latin

L'anglais n'est pas la langue la plus parlée au monde, mais c'est en revanche – et de loin – celle dont la diffusion géographique est la plus vaste et qui constitue la plus importante langue de communication internationale : l'anglais possède un statut officiel dans plus de 40 pays du monde ; 67% environ de l'ensemble des publications scientifiques sont rédigées en langue anglaise, sa présence sur internet est massive, etc.

Titicaca, de part et d'autre de la frontière entre le Pérou et la Bolivie. Aujourd'hui, leur langue, menacée par l'espagnol, survit assez bien en Bolivie où les Eglises catholique et évangélique publient plusieurs dizaines de périodiques en aymara. On estime que 500 000 Aymaras sont capables de lire leur langue. La situation est moins favorable au Pérou.

Typologiquement assez proche du quechua (la langue des Incas), l'aymara est typiquement « agglutinant », c'est-à-dire que sa grammaire comprend un grand nombre de suffixes invariables et souvent cumulables. Parmi ces nombreux suffixes, on trouve par exemple la marque du pluriel des noms (*-naka*) et celle de la négation (*-kiti*).

## Basque – euskara

### Famille

langue isolée (n'appartient pas au groupe des langues indo-européennes)

### Localisation

Pays Basque, une partie de la province de Navarre, France (Pyrénées-Atlantiques)

### Nombre de locuteurs

environ 600 000

### Système d'écriture

alphabet latin

Le basque est, en Europe occidentale, le seul vestige – très ancien – des langues parlées avant l'arrivée des Indo-européens.

Aujourd'hui langue co-officielle, avec le castillan, de la communauté autonome du Pays Basque (Espagne), le basque a été sévèrement réprimé depuis le XIX<sup>e</sup> siècle jusqu'à la fin du régime franquiste, qui était allé jusqu'à en interdire l'usage en Espagne, faire brûler publiquement les livres écrits en cette langue et supprimer les noms basques de la toponymie et des registres d'état civil! La situation juridique a considérablement changé depuis le retour de la démocratie: la province s'est vue accordée un statut d'autonomie, l'enseignement de la langue basque ainsi que des programmes scolaires bilingues ont été développés, etc. En 1982, le Parlement basque a adopté la *Loi fondamentale de normalisation de l'utilisation de l'euskara* (nom autochtone du basque), qui proclame le caractère co-officiel du castillan et du basque. L'usage du basque reste toutefois minoritaire (environ 20 à 25% de la population) et fortement soumis à la pression du castillan. Les services demeurent encore peu étendus, bien que les communications officielles écrites de l'administration se fassent presque toujours dans les deux langues, qu'il existe des programmes de radio et de télévision en basque, de nombreux journaux et magazines, etc.

La présence du basque en France remonte à la fin du VI<sup>e</sup> siècle, lorsque des groupes de Vascons espagnols émigrèrent vers le nord en traversant les Pyrénées jusqu'en Aquitaine (France) – qui fut alors connue sous le nom de Gascogne (par altération du mot espagnol *Vascones* ou du français *Vascons*, et désignant les ancêtres des Basques). Cependant, même si leur langue est aujourd'hui revalorisée et enseignée dans certaines écoles, les bascophones français ne bénéficient d'aucun véritable statut pour leur langue.

La langue basque se compose de plusieurs variétés dialectales (guipuzcoan, biscayen et haut-navarrais en Espagne, labourdin et bas-navarrais en France). L'intercompréhension peut être relativement difficile mais elle reste possible entre les différentes variétés. Depuis une trentaine d'années, un basque normalisé, unifié, a été adopté, essentiellement à l'écrit, pour remplir certaines fonctions sociales.

Le basque compte une dizaine de cas, 3 nombres (singulier, pluriel et « indéterminé ») et de très nombreux suffixes. Son système verbal, riche et complexe, contient 6 modes, 4 voix et des formes différentes selon qu'on s'adresse à un homme ou à une femme... En outre, le système de numération fonctionne en base 20 (« vicésimal »). Cependant, une part importante du vocabulaire provient de ses langues voisines: gaulois, latin, espagnol, français.

## Bengali

বাংলা

### Famille

langue indienne (famille indo-européenne)

### Localisation

langue officielle du Bangladesh et en Inde (Etats du Bengale et de l'Assam). D'importantes communautés migrantes le parlent en tant que langue première aux USA et en Grande-Bretagne.

### Nombre de locuteurs

207 millions

### Système d'écriture

alphabet bengali développé à partir de l'alphabet devanagari (système original du sanskrit)

Le bengali est la langue officielle du Bangladesh et l'une des seize langues « constitutionnelles » de l'Inde. Il possède une longue tradition littéraire. Son vocabulaire s'est enrichi de nombreux mots persans à l'époque où le persan a été langue officielle et administrative (domination des rois Moghols, fin du XVI<sup>e</sup> siècle) et, plus récemment, de nombreux termes anglais, surtout dans le domaine technique.

Le bengali connaît d'importantes différences dialectales. Phonétiquement, il est proche des autres langues indo-européennes de l'Inde. L'intercompréhension avec ces autres langues n'est pas automatique, mais elle ne pose pas de problèmes insurmontables, car leur unité est évidente (comparable selon Malherbe à celle qui lie les langues romanes).

## Berbères (langues)

### Famille

langues chamitiques (du groupe chamito-sémitique)

### Localisation

Maroc, Algérie, Lybie, Niger, ouest de l'Egypte, Mali

### Nombre de locuteurs

environ 15 millions

### Système d'écriture

alphabet *tifinagh* (uniquement pour certaines inscriptions)

Sous ce nom sont regroupées une vingtaine de langues, parlées par plus de 15 millions de Berbères. Le *kabyle*, par exemple, est parlé par plus de 3 millions de personnes à l'est d'Alger, le *chaoui* dans le massif des Aurès du Sud-est algérien (1,5 millions), le *tamazight* dans le nord et le centre du Maroc (3,5 millions), le *tashelhet* au sud du Maroc (3,5 millions), le *chleuh* au Maroc (1,5 million). Le *tamasheq*, parlé par environ 270 000 personnes, est la langue des Touaregs du Sahara. L'homogénéité des langues berbères est grande et leur parenté est aussi étroite qu'entre les langues latines les plus proches.

Historiquement, les langues berbères recouvraient une grande partie de l'Afrique du Nord, mais elles furent repoussées ou disparurent



suite au développement des cultures arabes. Seules ont subsisté les langues des populations des régions montagneuses inaccessibles et des déserts. Toutefois, soumises à une forte pression (liée à l'islamisation déjà très ancienne et favorisée par l'absence de culture écrite berbère) qui vise à arabiser les populations, souvent réprimées, les langues berbères peinent toujours à être reconnues officiellement (même si la situation semble aujourd'hui évoluer plus favorablement en Algérie pour la minorité kabyle). Il existe pourtant une ancienne écriture, le *tifinagh*, utilisée par quelques milliers de Berbères (surtout les Touaregs), mais elle n'est pas enseignée dans les écoles.

## Bosniaque

(voir chapitre sur le serbo-croate)

## Breton – brezhoneg

### Famille

langue celtique (famille indo-européenne)

### Localisation

France (Bretagne)

### Nombre de locuteurs

environ 250 000

### Système d'écriture

alphabet latin

Les langues celtiques constituent une branche de la famille indo-européenne. Elles ne sont plus parlées aujourd'hui qu'au pays de Galles, en Ecosse, en Irlande et en Bretagne par une population qui atteint environ 1 million et demi de personnes au total.

Le breton, peu écrit jusqu'à récemment, comporte d'importantes variations dialectales. Aujourd'hui, tous les Bretons parlent également le français et, sous la pression de l'école, des médias, du tourisme, leur langue est menacée. Elle a d'ailleurs pendant longtemps été l'objet d'un combat virulent de l'autorité centrale, en vue de l'éradiquer comme tous les «patois» de France. L'école a joué un rôle important dans ce travail d'éradication qui était réalisé au moyen de brimades, de punitions et d'interdictions («*Interdiction de cracher par terre et de parler breton*...»). Depuis lors, la situation a tout de même évolué. Aujourd'hui, 6000 élèves environ suivent un enseignement bilingue dans les écoles «*Diwan*». Et, sous l'impulsion de la *Délégation à la langue française et aux langues de France*, un important travail de description et de revalorisation des nombreuses langues recensées sur le territoire national a été entrepris.

Le breton est donc une langue parente du gaulois de l'Antiquité. Mais il n'en est pas le descendant direct, car la Bretagne a en fait été repeuplée aux V<sup>e</sup> et VI<sup>e</sup> siècles après J.C. par d'autres Celtes venus de Cornouailles. Et, contrairement à ce que l'on pourrait penser, le français compte peu de mots d'origine gauloise: seuls une septantaine de mots comme *bec*, *boue*, *chemin*, *mouton*, *alouette*... En revanche, les noms propres de villes, de lieux d'origine gauloise sont innombrables. En outre, les traces de numération vicésimale que conserve le français (trois vingts, six vingts jusqu'au XVIII<sup>e</sup> siècle... quatre-vingts aujourd'hui encore) proviendraient vraisemblablement d'un fond celtique.

### Sources

Walter, H. (1988). *Le français dans tous les sens*. Paris, Laffont.

## Bulgare – *bulgarski*

български

### Famille

langue slave (famille indo-européenne)

### Localisation

Bulgarie

### Nombre de locuteurs

environ 8 millions

### Système d'écriture

alphabet cyrillique

Langue slave la plus méridionale, le bulgare est parlé par 90% de la population de Bulgarie. Cette langue a joué un rôle important dans le développement de la famille des langues slaves: historiquement, en effet, c'est en vieux-bulgare que sont écrits les plus anciens textes slaves conservés jusqu'à nos jours (fin du X<sup>e</sup> siècle). Ainsi, le vieux-bulgare servit pendant longtemps de langue écrite de référence pour toutes les langues slaves. Et, au Moyen Age, il était l'une des trois langues écrites les plus importantes d'Europe. Il est en outre à l'origine du slavon qui fut la langue liturgique du rite byzantin.

Le macédonien est parfois considéré (surtout par les Bulgares!) comme une variété dialectale du bulgare.

## Cantonais (voir chinois)

Le cantonais est parlé par environ 6% des locuteurs de langue chinoise, dans le Guangdong, la partie sud du Guangxi, à Hong Kong et à Macao. C'est surtout l'une des principales langues de la migration: le 90% des Chinois vivant à l'étranger viennent des trois provinces du Sud-Est que sont le Guangdong, le Fujian et l'île de Hainan. C'est le seul dialecte chinois à avoir gardé la série complète des tons (il y en a neuf) ainsi que certaines autres particularités du chinois classique ancien. Le cantonais s'écrit comme le mandarin, avec des variantes locales. Il comporte quelques mots qu'on ne retrouve pas dans la langue commune et il a son propre système phonétique.

## Catalan – *català*

### Famille

langue romane (famille indo-européenne)

### Localisation

parlé au nord-est de l'Espagne, dans les Iles Baléares, à Andorre, dans les Pyrénées-Orientales françaises, en Sardaigne (Alghero)

### Nombre de locuteurs

environ 7 millions

### Système d'écriture

alphabet latin

Le catalan est une des langues officielles d'Espagne. Parlé, écrit et enseigné dans les régions autonomes du Nord-Est (Catalogne, Pays Valencien, Iles Baléares), il est également connu d'un nombre croissant de castillophones établis dans ces régions. En Catalogne, une très grande partie de la presse est en catalan et les écoles ont l'obligation d'offrir des enseignements dans cette langue. Ainsi, le bilinguisme catalan-castillan est de plus en plus répandu.

Le catalan a connu un important essor au Moyen Age, lorsqu'il était la langue du Royaume d'Aragon. Puis, après un relatif déclin et une renaissance au XIX<sup>e</sup> siècle, il a été interdit et violemment combattu

sous le régime franquiste (les livres en catalan n'existaient plus guère, car le possesseur d'un livre en catalan pouvait encourir l'emprisonnement !). En 1978, la nouvelle Constitution espagnole accorda à la Catalogne son autonomie politique et rétablit la *Generalitat de Catalunya*. Depuis lors, la Catalogne a soutenu avec un grand succès l'aménagement et le développement de sa langue par une politique volontariste : processus de *normalisation* du catalan, proclamation du catalan comme première langue de la Catalogne, nombreuses lois linguistiques à propos de l'affichage, de l'enseignement, etc.

Dans les autres régions catalanophones, qui sont elles aussi reconnues aujourd'hui par l'Etat espagnol, des variétés du catalan (*valencien, minorquin...*) ont également repris une relative importance, mais moindre tout de même que dans la riche et puissante Catalogne.

La langue catalane se rapproche en certains points des langues ibéro-romanes comme le castillan, en d'autres points des langues gallo-romanes, de l'occitan en particulier. Ainsi, le vocabulaire y est souvent plus proche du français que de l'espagnol (comme se dit *poma...*). Au niveau graphonétique on signalera entre autres les graphies *ll*, prononcées [lj], et *ig*, prononcées [tj].

## \* Chinois (voir aussi cantonais) 中文

### Famille

langue sino-tibétaine

### Localisation

République populaire de Chine, Taïwan, Singapour

### Nombre de locuteurs

plus d'un milliard quatre cents millions

### Système d'écriture

écriture chinoise, de type idéographique

### Les langues en Chine

En Chine même, le chinois ou « langue des Han », du nom de l'ethnie majoritaire représentant 94 % de la population totale, n'est que l'une des 55 langues répertoriées officiellement : 6 % des locuteurs de nationalité chinoise, soit environ 55 millions de personnes, parlent une langue d'un autre groupe (turc, mongol, toungouse, coréen, tibéto-birman, thaï, miao ou yao).

### Les principaux dialectes chinois et la langue commune

Les Chinois parlent des dialectes aussi différents les uns des autres que le sont les diverses langues romanes d'Europe. On oppose une zone dialectale relativement homogène, qui domine au nord de la Chine et comprend le Sichuan ainsi que les populations chinoises du Yunnan et du Guizhou, à une zone de grande diversité de parlers, au sud-est. Ces dialectes du Sud-Est se divisent en six grands groupes : wu (Jiangsu), gan (Jiangxi), xiang (Hunan), cantonais (Guangdong), hakka (Guangdong, Fujian et Taïwan) et min (Fujian). La langue commune ou *putonghua*, communément appelée « mandarin », est la langue officielle basée sur le parler de la première zone dialectale. Le mandarin est obligatoire dans l'enseignement, l'administration et les médias nationaux.

Le chinois est très différent des langues indo-européennes que nous connaissons. Les mots sont monosyllabiques et invariables, de sorte qu'il n'y a ni conjugaisons, ni déclinaisons. Il n'y a en chinois ni nombre, ni genre, ni temps du verbe. Il n'y a pas non plus d'adjectifs à proprement parler, ni de prépositions. La grammaire ne comporte donc pas de partie morphologique et se ramène à la

seule dimension de la syntaxe. Cette syntaxe se réduit, en chinois moderne, au maniement d'une douzaine de particules, pas plus, et à l'application d'un nombre limité de règles concernant la position des mots dans la phrase. Le système est d'une grande simplicité mais c'est de là que vient, paradoxalement, la difficulté pour celui qui apprend cette langue : ce n'est pas du jour au lendemain que l'on apprend à redresser ses idées selon des relations plus simples que celles dont on a pris l'habitude depuis l'enfance...

En outre, le chinois est une langue *tonale*. Le premier ton (transcrit *mā*) est placé haut, il est égal et long, comme lorsqu'on fait « aaaaa » chez le médecin ; le 2<sup>e</sup> ton (transcrit *má*) est placé haut, il est ascendant et plus bref, correspondant à l'intonation de l'interrogation en français ; le 3<sup>e</sup> ton (*mǎ*) est placé bas, il est incurvé et long ; enfin, le 4<sup>e</sup> ton (*mà*) part de très haut, est descendant et très bref, comme lorsqu'on dit « non ! » de manière péremptoire. Alors qu'en français la mélodie donne uniquement une indication du type de la phrase (déclaratif, interrogatif, etc.), en chinois elle donne sens au mot : *mā* veut dire « maman », *má* « chanvre », *mǎ* « cheval » et *mà* « dire des gros mots ».

### L'écriture idéographique

L'écriture chinoise est dite « idéographique » en opposition à l'écriture alphabétique. Un mot chinois monosyllabique s'appelle un « caractère » (les Français utilisent volontiers le terme de « sinogramme »). L'écriture chinoise se compose de milliers de caractères qui correspondent chacun à un mot. On peut donner une première idée de sa fonction en l'assimilant à nos chiffres : 5 désigne un nombre sans préjuger de la manière dont un Français, un Italien ou un Espagnol appelleront ce nombre dans leurs parlers respectifs – *cing, cinque, cinco*, etc. De même, le caractère chinois désigne un mot ou l'idée d'un mot (son signifié) sans préjuger de la manière dont le mot est prononcé par des Chinois originaires de différentes époques ou parties de la Chine et parlant des dialectes différents.

Pour faire correspondre un caractère à chaque mot, il a fallu créer des milliers, voire des dizaines de milliers de caractères. Les Chinois ont non seulement su créer un vocabulaire graphique d'une grande richesse, mais ils ont su faire en sorte qu'il reste assimilable, c'est-à-dire qu'il n'exige pas de la mémoire un effort démesuré. Ils y sont parvenus grâce à la combinaison souple de plusieurs procédés. En simplifiant, on peut classer les caractères en quatre catégories fondamentales.

On classe dans la 1<sup>re</sup> catégorie les plus anciens caractères chinois, des dessins simplifiés d'objets, d'êtres vivants ou de phénomènes naturels devenus signes conventionnels, autrement dit des *pictogrammes*. Il n'y en a pas beaucoup, environ 360. Les caractères classés dans la 2<sup>e</sup> catégorie expriment des idées plus abstraites : leur signification est suggérée par une idée ou par l'association de deux idées. Ce sont les *idéogrammes*, dont le procédé a permis d'augmenter considérablement le nombre des caractères, jusqu'à plus de 1000. Les caractères de la 3<sup>e</sup> catégorie, dits « à forme et à son », sont composés d'un élément pictographique simple (appelé « clé » ou « radical ») donnant une indication de sens et d'un autre caractère (simple ou composé) pris pour sa valeur phonétique et qui fournit une indication de son. Ce sont les *idéophonogrammes*, un procédé qui a permis de multiplier rapidement le nombre des caractères et de former des milliers de composés facilement assimilables. Une 4<sup>e</sup> catégorie concerne les *emprunts* : on parle de caractère emprunté lorsqu'un même caractère est utilisé pour noter deux mots homophones qui n'ont aucun rapport entre eux du point de vue du sens.



Le caractère chinois s'écrit en équilibre au centre d'un carré tracé ou imaginaire selon un certain nombre de traits plus ou moins grand. L'ordre des traits est contraignant et il faut suivre des règles strictes pour des questions de lisibilité. Les carrés dans lesquels s'inscrivent les caractères peuvent être juxtaposés sans inconvénient dans l'un ou l'autre sens : on écrit de droite à gauche en colonne, mais de gauche à droite en ligne.

#### La transcription phonétique pinyin

A la fin des années 50, la Chine officialisait une transcription phonétique en lettres latines du mandarin, dite pinyin. Cette translittération se généralise depuis les années 90 dans les publications occidentales. Le pinyin est utilisé dans l'enseignement élémentaire et dans les centres d'enseignement du chinois comme langue étrangère, mais peu d'adultes chinois le maîtrisent vraiment.

Pour des raisons de commodité, nous utilisons ici, pour transcrire le chinois, le même système que pour les autres langues dont le système d'écriture diffère de l'alphabet latin (cf. *Quelques indications pour la lecture et la prononciation des langues*).

## Corse – corsu

### Famille

langue romane de la famille indo-européenne (apparentée au groupe italien)

### Localisation

Corse

### Nombre de locuteurs

environ 300 000

### Système d'écriture

alphabet latin

La Corse est devenue française en 1796. Auparavant, elle fut colonie romaine, puis elle fut dominée successivement par Pise (XI<sup>e</sup>-XIII<sup>e</sup> siècle) et par Gênes pendant cinq siècles. Jusqu'au XIX<sup>e</sup> siècle, toute la population parlait le corse, tandis que la classe cultivée parlait italien. Aujourd'hui, le bilinguisme français-corse s'est généralisé, mais l'idiome de l'île est resté une langue très vivante, aussi bien à la campagne qu'en ville. Contrairement à ce qui s'est passé dans d'autres régions de France, l'usage généralisé du français n'a pas fait reculer celui du corse et l'attachement des Corses à leur langue reste exemplaire.

Apparenté à l'italien (plus particulièrement au toscan), le corse a conservé certains traits anciens du latin disparus de l'italien, comme par exemple la voyelle *u* (prononcée [u]) en position finale (« mur » se dit *mur* en corse mais *muro* en italien, « année » se dit *annu* en corse mais *anno* en italien, etc.). Il existe en outre une grande diversité de variétés locales de la langue, mais une graphie unique a été adoptée, basée sur l'orthographe italienne (*gh* se prononce [g], *ch* se prononce [k], *c* se prononce [tʃ], etc.).

### Tiré de

Walter, H. (1994). *L'Aventure des langues en Occident*. Paris, Laffont, p. 292.

Walter, H. (1998). *Le français dans tous les sens*. Paris, Laffont, p. 154.

## Créoles (langues)

### Famille

voir descriptif

### Localisation

Les créoles se trouvent essentiellement dans les anciennes zones d'esclavage (Caraïbes et Océan Indien), mais on en trouve aussi qui sont apparus de façon atypique en Afrique même.

- *Créoles français* : Antilles françaises (Guadeloupe, Martinique), Haïti, certaines petites Antilles ex-britanniques (Sainte-Lucie, Dominique), Guyane française, Brésil (Etat d'Amapa), île de la Réunion, île Maurice, îles Seychelles.
- *Créoles anglais* : Jamaïque, petites Antilles, Surinam et Guyane française (sranan tongo et créoles noirs marrons), Guyana.
- *Créoles portugais* : Antilles néerlandaises, Afrique lusophone (Cap Vert, Guinée-Bissau)
- *Créole espagnol* : Colombie
- *Créole néerlandais* : Guyana

### Nombre de locuteurs

plusieurs millions pour les créoles français et anglais ; nombre très restreint pour les créoles espagnol et néerlandais

### Système d'écriture

alphabet latin.

On appelle créoles des langues apparues dans des conditions « de contact », et généralement de contact brutal, le cas de figure prototypique étant celui de l'esclavage dans la zone caraïbe (insulaire ou continentale) et dans l'océan Indien. Sur les plantations étaient réunis des colons européens et des esclaves d'origine africaine qui ne parlaient pas la même langue et étaient donc obligés de recourir à la langue des maîtres pour communiquer entre eux. Mais cette langue était acquise dans des conditions défavorables (conditions matérielles et morales misérables, transmission d'une langue restreinte par des ordres du maître ou des contremaîtres esclaves de première génération). Il y a donc eu un processus de destruction de la langue européenne suivi d'un processus d'appropriation et de restructuration par les esclaves. Cette apparition de nouvelles langues s'est en général produite vers le milieu du XVII<sup>e</sup> siècle et a été très rapide (en une génération un parler créole peut devenir langue maternelle). Toutes les langues créoles présentent les deux caractéristiques suivantes :

1. Le vocabulaire est massivement (90% voire plus) issu d'une langue de colonisation : il existe ainsi des créoles dits anglais, français, espagnols, portugais, néerlandais.
2. La grammaire est totalement restructurée, en particulier en ce qui concerne le système verbal et la détermination nominale, et dans des formes qui se ressemblent beaucoup d'un créole à l'autre (mais ne ressemblent pas à la grammaire des langues européennes).

## Croate

(voir chapitre sur le serbo-croate)

## Danois – deens

### Famille

langue germanique (branche scandinave) de la famille indo-européenne

### Localisation

Danemark, Groenland

### Nombre de locuteurs

environ 5 millions

**Système d'écriture**

alphabet latin

Le danois est très proche du suédois et encore plus du norvégien. Il a en effet longtemps été utilisé par les Norvégiens lettrés à l'époque où les Danois administraient leur pays. Le danois est avec le *kalaallisut* la langue co-officielle du Groenland (*Grønland*, pays vert), département danois jouissant d'une large autonomie; il y est parlé par la minorité danoise (dix à quinze mille personnes) et, comme langue seconde, dans diverses situations formelles.

Le danois possède deux genres: un genre «commun» et un «neutre», le féminin ayant disparu. Le système graphique comporte quelques particularités ( $\emptyset = [\emptyset]$ ,  $\text{æ} = [\text{ɛ}]$ ,  $\text{å} = [\text{o}]$ ...). En 1948, une réforme a aboli l'emploi d'une majuscule à l'initiale des noms communs.

**\* Espagnol – *español* (ou *castellano*)****Famille**

langue romane (famille indo-européenne)

**Localisation**

Espagne, tous les pays d'Amérique latine, à l'exception du Brésil et du Belize; Guinée Equatoriale (où il est langue officielle). L'espagnol est aussi très fortement présent aux USA.

**Nombre de locuteurs**

plus de 300 millions

**Système d'écriture**

alphabet latin

L'espagnol est une langue internationale majeure et l'une des six langues des Nations Unies. Il a un statut officiel dans dix-huit pays. L'unité de l'espagnol parlé en Espagne et dans les pays d'Amérique latine est remarquable, même s'il existe toutefois des variations régionales, en particulier dans le vocabulaire. L'arabe a eu une influence profonde sur la langue en raison de la présence arabe sur une partie de la péninsule ibérique entre 711 et 1492: on recense en effet près de 4000 mots d'origine arabe en espagnol (la plupart des mots commençant par *al-* sont d'origine arabe: *almohada*, coussin; *alfombra*, tapis; *algodón*, coton, ...). En français, près de 300 mots d'origine étrangère sont passés par l'espagnol: il peut s'agir de mots antérieurement arabes ou de mots désignant des plantes ou des animaux originaires des Amériques.

L'espagnol comporte de nombreuses variantes dialectales dont l'andalou, le parler de la Manche, celui des Asturies, etc. (variétés qui ne doivent pas être confondues avec les autres langues parlées sur le territoire espagnol: basque, galicien, catalan).

L'écriture de l'espagnol a la particularité de commencer les phrases interrogatives par un point d'interrogation écrit à l'envers (idem pour les exclamatives).

**Espéranto – *esperanto*  
(la *lingvo internacia*)****Famille**

langue artificielle inventée (au XIX<sup>e</sup> siècle) dont le vocabulaire est largement d'origine latine (plus de 80%) ou d'autres langues indo-européennes (grec, langues germaniques,...) mais dont la grammaire est bien différente.

**Localisation**

parlé dans plus de 115 pays, plutôt en Europe centrale et de l'Est, mais aussi en Chine et d'autres pays d'Asie et dans quelques régions d'Amérique du Sud.

**Nombre de locuteurs**

200-2000 en tant que langue première; environ 2 millions de locuteurs comme deuxième langue.

**Système d'écriture**

alphabet latin

Langue de synthèse, langue artificielle, l'espéranto a été créé par L.-L. Zamenhof, médecin et linguiste polonais polyglotte, à la fin du XIX<sup>e</sup> siècle, dans le but de faciliter la compréhension entre les peuples et de contribuer ainsi à promouvoir une coexistence pacifique et fraternelle entre les êtres. L'espéranto a donc un aspect idéologique et propédeutique. Il apparaît comme un pont entre les différentes langues, d'Europe en particulier. Les espérantistes tentent constamment d'augmenter la diffusion de leur langue au moyen de nombreux enseignements et de nombreuses traductions en espéranto. Il existe même aujourd'hui des locuteurs natifs – élevés dans cette langue – et des œuvres rédigées directement dans cette langue. Mais s'il connaît de nombreux adeptes, l'espéranto a aussi des détracteurs qui ne croient pas à la viabilité d'une telle langue artificielle et lui reprochent ses origines trop indo-européennes.

C'est une langue aux règles très simples et sans exceptions, dont le lexique est construit à partir de racines courantes des langues occidentales (indo-européennes) les plus répandues. Les racines de base du vocabulaire peuvent donner naissance à quantité de mots composés et/ou dérivés grâce au jeu de préfixes (*mal-*, *re-*, etc.) et de suffixes (*-et* (= diminutif), *-ar* (= collectif), *-in* (= féminin), etc.).

**Finnois – *suomi*****Famille**

langue finno-ougrienne (ou ouralienne)

**Localisation**

Finlande, nord de la Suède et, dans une moindre mesure, en Russie

**Nombre de locuteurs**

environ 6 millions

**Système d'écriture**

alphabet latin

Le finnois est l'une des rares langues d'Europe à ne pas faire partie de la famille indo-européenne. Comme l'estonien et le hongrois, il fait partie de la famille finno-ougrienne (ou ouralienne) qui rassemble une population de plus de 25 millions de locuteurs se partageant une trentaine de langues. Son origine remonte à plusieurs siècles dans les régions situées à l'est de l'Oural (chaînes de montagnes qui séparent l'Europe de l'Asie, en Russie).

Le finnois est une langue agglutinante qui contient de nombreux suffixes. Il ne distingue pas le genre et n'a pas de déterminant. En revanche, il possède 15 cas différents pour les noms (*nominatif*, *accusatif*,... mais aussi *élatif* (d'où on vient), *essif* (état ou lieu dans lequel on se trouve), etc.), marqués par des désinences régulières qui se rapprochent de postpositions. L'alphabet finnois utilise 21 lettres et son système graphophonétique est très régulier. A l'exception des emprunts linguistiques, il n'existe aucune similarité entre les langues indo-européennes et le finnois, ce qui rend son apprentissage difficile.



## \*Français – français

### Famille

langue romane (famille indo-européenne)

### Localisation

langue officielle de France (et ses départements et territoires d'Outre-mer), Belgique, Canada, Luxembourg, Suisse, Italie (Val d'Aoste), Monaco, Bénin, Burkina Faso, Burundi, Cameroun, Comores, Congo, Côte d'Ivoire, Djibouti, Gabon, Guinée, Haïti, Mauritanie, Madagascar, Mali, Niger, Rwanda, République centrafricaine, République démocratique du Congo, Sénégal, Tchad, Togo.

### Nombre de locuteurs

77 millions

### Système d'écriture

alphabet latin

Le français représente l'une des langues les plus employées en tant que langue internationale de communication et dans le domaine de la diplomatie. Toutefois, elle n'occupe aujourd'hui dans cette fonction que le deuxième rang loin derrière l'anglais. Elle n'en reste pas moins utilisée comme langue d'usage, à des degrés divers, aussi bien en Europe que sur les autres continents :

- En *Afrique*, le français est langue officielle dans 18 pays (seule ou avec d'autres langues). Cependant, la grande majorité de la population n'est pas francophone. En revanche, dans les trois pays du Maghreb (Maroc, Tunisie et Algérie), le français n'est pas langue officielle, mais il est compris et parlé par une partie appréciable de la population.
- En *Amérique du Nord*, le français est parlé quotidiennement au Canada (Québec, Nouveau-Brunswick, Ontario) et dans certaines communautés francophones aux Etats-Unis (Missouri, Vermont, Maine... et, surtout, Louisiane) où il n'a toutefois pas le statut de langue officielle.
- Dans *les Antilles*, outre les Antilles françaises (Martinique, Guadeloupe et dépendances), la République de Haïti a gardé le français comme seule langue officielle.
- En *Amérique du Sud*, la Guyane française est devenue un département d'outre-mer en 1946, après avoir été une colonie depuis le milieu du XIX<sup>e</sup> siècle.
- Au *Moyen-Orient*, le Liban occupe une place spéciale : le français y conserve un réel prestige et semble rester la langue de culture, même si l'anglais gagne du terrain depuis une vingtaine d'années.
- Dans *l'Océan Indien*, le français est resté la langue de l'élite intellectuelle de Madagascar longtemps après l'indépendance. Elle y est aujourd'hui encore langue officielle avec le malgache. Les autres îles francophones sont territoire ou département français d'outre-mer (Mayotte et l'île de la Réunion).
- En *Asie*, le français fut une langue bien installée dans l'ancienne Indochine française où un bilinguisme authentique existait chez une partie de la population. On observe encore un certain maintien du français au Laos et au Cambodge.
- En *Océanie*, le français est présent dans trois territoires d'outre-mer, la Nouvelle-Calédonie, les îles Wallis et Futuna, et la Polynésie française.

Parmi les langues du monde, le français se distingue entre autre par les enjeux politiques auxquels il a toujours été soumis dès sa recon-

naissance officielle comme langue du Royaume ; en 1539, dans l'ordonnance de Villers-Cotterêts, François I<sup>er</sup> décrète que les textes juridiques seront désormais rédigés en « langage maternel français ». Le français a dès lors été utilisé et perçu comme l'expression de l'unification nationale, surtout après la révolution qui voulait en faire un instrument de l'universalisme et de l'égalité républicaine. Il en est résulté une politique linguistique – illustrée par exemple par l'Académie française – au caractère fortement centralisateur et normatif qui a largement contribué, au travers d'une véritable politique d'éradication, au recul des dialectes et autres langues en usage sur le territoire national et – bien souvent – à la culpabilisation linguistique (orthographique en particulier) des locuteurs...

On rappellera pourtant que le français connaît aujourd'hui encore, à côté de la langue normée, des variétés dialectales importantes (*poitevin, tourangeau, picard, franc-comtois...* en France, *wallon* en Belgique, *suisse romand...*). Ces variétés remontent aux trois groupes dialectaux en usage au Moyen Age : dialectes d'oc, d'oïl et franco-provençaux (auxquels appartiennent les patois de Suisse romande, à l'exception du patois jurassien qui relève des dialectes d'oïl). De plus, le français a connu une nouvelle diversification à travers son expansion dans le monde, en Afrique, au Québec, dans les îles Caraïbes, etc. Certains des pays et régions qui l'utilisent se risquent d'ailleurs aujourd'hui à revendiquer une certaine autonomie par rapport aux normes de la France (comme en témoignent par exemple l'élaboration de dictionnaires de français propres à ces pays ou régions) et un droit de regard sur la gestion commune de la langue.

### Source

Walter, H. (1988). *Le français dans tous les sens*. Paris, Laffont.

## Galicien – galego

### Famille

langue romane (famille indo-européenne)

### Localisation

Galice (Nord-Ouest de l'Espagne)

### Nombre de locuteurs

environ 3 millions

### Système d'écriture

alphabet latin

Le galicien est une langue romane proche du portugais. Jusqu'au XII<sup>e</sup> siècle, il s'agissait d'ailleurs d'une même langue qui s'est différenciée par la suite. Aujourd'hui, la population de Galice s'élève à environ 3 millions d'habitants, dont quelques 72% parlent le galicien comme langue maternelle. Tous les Galiciens connaissent toutefois le castillan. Le galicien est également parlé par un million de Galiciens émigrés en Amérique ou en Europe.

Après la domination franquiste, marquée par l'interdiction de la langue régionale, le statut d'autonomie accordé en 1981 à la Galice lui a permis d'adopter des lois linguistiques en conformité avec la constitution de l'Etat espagnol. Ainsi, selon l'article 4 de la Loi sur la normalisation linguistique (1983) : « le galicien en tant que langue propre de la Galice, est la langue officielle des institutions de la Communauté autonome, de son administration, de l'administration locale et des entités publiques dépendant de la Communauté autonome ». Cependant, la « reconquête » linguistique est plus difficile qu'en Catalogne par exemple, d'une part du fait que la langue régionale galicienne jouit d'un prestige moindre, d'autre part du fait que les autorités ne développent pas une politique linguistique aussi

volontariste. Ainsi, en Galice coexistent aujourd'hui deux formes de galicien correspondant à des orientations différentes: d'un côté le «galicien historique», resté très proche du portugais moderne et aisément compréhensible, à l'écrit comme à l'oral, par les lusophones; de l'autre côté une variété désignée aujourd'hui par le terme de *castrapo* ou par l'expression *galego espanholizado* (galicien hispanisé), fortement influencée par le castillan.

## Grec

ἑλληνικά

### Famille

langue «isolée» de la famille indo-européenne

### Localisation

Grèce, Chypre, Albanie, Egypte, Italie (quelques villages en Italie du Sud)

### Nombre de locuteurs

environ 12 millions

### Système d'écriture

alphabet grec

Le grec est une langue aujourd'hui isolée, d'une longévité et d'une continuité pourtant remarquables et qui a connu un extraordinaire rayonnement culturel. Son impact sur les autres langues a en effet été immense, d'abord à travers son système d'écriture, mais aussi d'un point de vue lexical: des préfixes tels que *poly-*, *micro-*, *anti-*, *héli-*, *hétéro-*, *auto-*, *chrono-*, etc., des suffixes tels que *-logie*, *-chrome*, *-phile*, etc. ont généré d'innombrables mots du langage scientifique ou courant.

Le grec ancien n'est plus parlé aujourd'hui, excepté comme langue religieuse dans l'église grecque orthodoxe. Il fut également la langue officielle de l'Empire Byzantin du IV<sup>e</sup> au XV<sup>e</sup> siècle. Le grec moderne (démotique) prit forme aux environs du IX<sup>e</sup> siècle et devint la langue officielle dans le royaume grec au XIX<sup>e</sup> siècle. Il possède trois genres (masc., fém., neutre). En revanche, il a abandonné le «duel» du grec ancien et n'a conservé que quatre cas.

L'alphabet grec, adapté de l'alphabet phénicien, existe depuis environ 1000 ans avant notre ère. C'est le premier alphabet qui introduisit des lettres pour marquer les voyelles, se différenciant ainsi des alphabets sémitiques qui ne notaient jusqu'alors que des consonnes. Comme les alphabets sémitiques, il s'écrivait à l'origine de droite à gauche, puis il s'est écrit dans un style où les lignes allaient alternativement de droite à gauche puis de gauche à droite, avant, finalement, de s'écrire uniquement de gauche à droite. Aujourd'hui, les lettres grecques sont largement utilisées en mathématique. L'alphabet du grec moderne est resté le même mais la prononciation de certaines voyelles s'est simplifiée. On notera que les lettres  $\theta$  et  $\delta$ , utilisées dans les mots de certaines activités, correspondent aux «th» anglais de *thing*, respectivement *this*.

## Gujrati

ગુજરાતી

### Famille

langue indo-européenne de l'Inde (restée proche du sanskrit)

### Localisation

Inde (états de l'Ouest: Gujarat, Maharashtra); communautés importantes en Afrique de l'Ouest (Tanzanie, Ouganda...), en Grande-Bretagne et en Australie

### Nombre de locuteurs

environ 45 millions

### Système d'écriture

alphabet gujrati (proche de celui de l'hindi)

Le gujrati est l'une des 16 langues constitutionnelles de l'Inde et possède le statut de langue officielle de l'Etat du Gujarat (ouest), où il est parlé par plus de 80% de la population. Langue préférée du Mahatma Gandhi, originaire de cet Etat, le gujrati connaît des variations dialectales portant sur le lexique et liées à l'appartenance religieuse des diverses communautés qui l'utilisent: hindouistes, musulmans, ismaéliens, parsis (zoroastriens).

Phonétiquement et grammaticalement, le gujrati est proche des autres langues indo-européennes de l'Inde (*cf.* hindi). L'intercompréhension avec ces autres langues n'est pas automatique, mais ne pose pas de problèmes importants.

## Haoussa – hausa

### Famille

langue du groupe tchadique (parfois rapproché des langues chamito-sémitiques)

### Localisation

Nigeria, Niger, Soudan

### Nombre de locuteurs

environ 24 millions en tant que langue première, environ 39 millions en tant que langue première ou seconde

### Système d'écriture

alphabet latin

Le haoussa est la langue d'Afrique noire la plus parlée. Il est l'une des 8 langues officielles du Niger et l'une des langues majeures du Nigeria, à côté de l'*ibo*, du *yorouba* et de l'*anglais* (langue officielle). Il est de plus utilisé par au moins 50% de la population du Nigeria en tant que langue seconde et il jouit d'un rayonnement culturel important, entre autre comme langue véhiculaire des commerçants, en usage jusqu'à l'Afrique équatoriale.

Parmi les caractéristiques originales de la grammaire haoussa, on notera l'existence de deux genres (féminin et masculin) nettement marqués et l'absence de classe nominales, contrairement à ce que l'on trouve dans les langues de la famille bantoues. Le haoussa a été profondément imprégné par l'influence de l'islam, l'arabe ayant transmis un important vocabulaire dans tous les domaines. Les langues européennes, en particulier l'anglais, sont à l'origine du vocabulaire technique ou moderne: *asibiti* (*hospital*), *fasofis* (*post-office*).

## Hébreu

עברית

### Famille

langue sémitique

### Localisation

Israël

### Nombre de locuteurs

environ 5,2 millions

### Système d'écriture

alphabet hébreu (s'écrit de droite à gauche)

L'hébreu moderne présente cette particularité de n'être pas directement issu de l'ancien hébreu (dont les premiers textes bibliques remontent à environ 1000 ans av. J.-C.). Il a en effet été réactivé,



revivifié dès la fin du XIX<sup>e</sup> siècle, après une longue période durant laquelle il n'existait plus que comme langue religieuse; de nouveaux mots ont ainsi été créés à partir d'anciennes racines hébraïques, mais aussi arabes et araméennes.

L'hébreu est aujourd'hui la langue principale de l'Etat d'Israël, utilisée à des degrés divers par près de 80% de la population juive et arabe du pays. Mais il n'est langue maternelle que pour moins de 50% de la population (juive).

L'hébreu est clairement une langue sémitique (terme créé par les linguistes, à partir de Sem, nom d'un des fils de Noé), comme l'*arabe*, l'*amharique*, etc., avec lesquels la parenté lexicale et grammaticale est bien visible: mots formés à partir de radicaux à trois consonnes, présence de consonnes gutturales, etc.

Son alphabet, développé à partir de l'alphabet phénicien, s'écrit de droite à gauche et ne note pas les voyelles courtes (il existe toutefois un système de notation, au moyen de points, utilisé pour l'apprentissage).

On notera que le *yiddish*, qui a constitué depuis le X<sup>e</sup> siècle la *lingua franca* de la communauté juive d'Europe centrale et orientale (dite ashkénaze), est une langue germanique (80% du lexique), mais qui a également beaucoup emprunté aux langues slaves et à l'hébreu (la formation du pluriel, par exemple, est la même qu'en hébreu). En outre, l'écriture du yiddish est fondée sur l'alphabet hébraïque.

## \* Hindi

हिन्दी

### Famille

langue indo-européenne de l'Inde, très proche de l'ourdou

### Localisation

Nord de l'Inde, Népal (Afrique du Sud et de l'Ouest)

### Nombre de locuteurs

environ 337 millions

### Système d'écriture

alphabet devanagari (utilisé à l'origine pour le sanskrit)

L'hindi est, avec l'anglais, l'une des deux langues officielles nationales de l'Inde. Parmi les nombreuses langues de ce pays, c'est donc la plus importante. Cependant, elle reste pour la plupart des Indiens une langue «étrangère», perçue comme la langue des castes hindoues dirigeantes, que l'Etat a tenté d'imposer à l'ensemble du pays. Nombre d'entre eux, surtout ceux du Sud qui utilisent une langue dravidienne, se sont opposés à la prépondérance de l'hindi et ont préféré utiliser l'anglais, langue «moins conflictuelle». Cette situation explique en partie son rayonnement international très relatif.

Historiquement, l'hindi est très proche de l'ourdou parlé au Pakistan. Par volonté de différenciation, il a cependant choisi le système d'écriture devanagari et remplacé une partie de son vocabulaire arabisé par des termes repris du sanskrit. Aujourd'hui, l'hindi contient de nombreux emprunts à l'anglais.

Grammaticalement, l'hindi possède deux genres (masculin – féminin); c'est une langue «à cas», qui en possède six: nominatif – accusatif – génitif – datif – locatif – instrumental, et qui contient en outre des postpositions (et non des prépositions). Le système de conjugaison, plutôt riche, contient entre autres des formes progressives de présent et de passé (cf. formes en *-ing* de l'anglais). La phonétique de l'hindi comporte quelques sons «expirés» étrangers au français et la longueur des voyelles joue un rôle important.

## Hollandais (voir néerlandais)

### Hongrois – magyar

#### Famille

langue finno-ougrienne (ou ouralienne)

#### Localisation

Hongrie, Roumanie (1,5 million), petites minorités en Ex-Yougoslavie et en Slovaquie

#### Nombre de locuteurs

environ 14 millions

#### Système d'écriture

alphabet latin

Le hongrois est l'une des rares langues d'Europe à ne pas faire partie de la famille indo-européenne. Comme l'estonien et le finnois, il fait partie de la famille finno-ougrienne (ou ouralienne) qui rassemble une population de plus de 25 millions de locuteurs se partageant une trentaine de langues. Il semble que le peuple hongrois, venu d'Asie, a vécu longtemps en tant que nomade dans la région de l'actuelle Ukraine avant de s'établir au IX<sup>e</sup> siècle sur les bords du Danube où il vit aujourd'hui.

Le hongrois est considéré comme étant difficile à apprendre en raison de son vocabulaire où l'on ne retrouve pas de racines indo-européennes. Toutefois, il contient de nombreux mots empruntés aux langues slaves, au latin (qui a été jusqu'au XVI<sup>e</sup> siècle langue officielle en Hongrie!), à l'allemand ainsi qu'au turc. Il s'agit d'une langue qui marque les fonctions par un vaste système de suffixes qui varient selon le contexte phonétique (mécanisme de l'harmonie vocalique caractéristique des langues finno-ougriennes, mais aussi du turc): a *házban* (dans la maison), a *kertben* (dans le jardin). Comme les autres langues finno-ougriennes, le hongrois ne distingue pas le genre des noms. L'alphabet contient 26 lettres; les relations graphie-phonie y sont tout à fait régulières.

### Indonésien – bahasa indonesia

#### Famille

langue malayo-polynésienne

#### Localisation

Indonésie

#### Nombre de locuteurs

environ 20 millions

#### Système d'écriture

alphabet latin

Depuis l'indépendance du pays en 1945, l'«indonésien commun» (*bahasa indonesia*) a été décrété langue nationale de ce pays de près de 200 millions d'habitants et qui compte plusieurs centaines de langues, la plupart malayo-polynésiennes (javanais, soundanais...). Mais il n'est parlé en Indonésie que par une petite minorité de 6,7 millions de Javanais et d'un autre groupe de 10 millions de Malais dans l'île de Sumatra: moins de 10% des Indonésiens parlent ainsi l'indonésien officiel comme langue maternelle.

On observe là un cas très intéressant d'aménagement linguistique où une langue minoritaire a été promue langue nationale avec le consentement de la plupart des locuteurs. Dès l'indépendance, cette langue a fait l'objet d'un travail de codification, de normalisation et

de diffusion planifiée (publication de grammaires, de dictionnaires et de manuels pédagogiques, création de terminologies scientifiques...). Elle a été simplifiée, afin de constituer une langue commune passe-partout, en étroite parenté avec les autres langues indonésiennes. Et elle est ainsi devenue pour 150 millions de personnes en Indonésie (c'est à dire plus de trois quarts de la population) une langue *seconde* qu'ils comprennent et, parfois, utilisent en complément à leur langue locale.

Linguistiquement, l'indonésien est par conséquent bien plus simple que les langues malayo-polynésiennes traditionnelles. Il ne possède par exemple plus d'« infixes », pas de flexion des noms, ni de conjugaison, les noms n'ont généralement pas de genre, et on ne marque le pluriel, par redoublement du mot, que si cela est nécessaire à la compréhension. En revanche, la formation des mots, par préfixation, suffixation ou composition, y est très riche. Le vocabulaire comporte de nombreux mots d'origine arabe, en raison de l'influence de l'islam, mais aussi portugaise et hollandaise, en liaison avec l'ancienne colonisation du pays. La prononciation ne pose pas de problème particulier.

## Islandais – *íslensk*

### Famille

langue germanique (branche scandinave) de la famille indo-européenne

### Localisation

Islande

### Nombre de locuteurs

environ 250 000

### Système d'écriture

alphabet latin complété par deux consonnes originales: le *ð* (*th* anglais de *that*) et le *þ* (proche du *th* anglais de *thin*)

En raison de son isolement géographique, l'islandais est resté très proche du vieux-norvégien, la langue des Vikings qui fut amenée en Islande au IX<sup>e</sup> siècle. C'est ainsi une langue conservatrice, voire « archaïque » et homogène, qui possède trois genres, quatre cas et qui, comme les autres langues scandinaves, place les articles définis après le nom.

Après une période d'indépendance, l'Islande a été longtemps sous la domination du Danemark qui a cherché à imposer sa langue. Ainsi, jusqu'au début du XX<sup>e</sup> siècle, l'islandais n'était pas enseigné dans les écoles; pourtant, il s'est remarquablement maintenu et, dès l'indépendance du pays, en 1944, il a été l'objet d'une politique linguistique active afin de lui permettre de fonctionner dans tous les domaines de la vie moderne (éducation, science, administration...). Une « Commission de la langue islandaise » est chargée de légiférer en ce domaine et, en particulier, d'adapter la terminologie en « islandisant » les termes techniques modernes; par exemple téléphone se dit *sími* (terme formé à partir d'un ancien mot signifiant « lien », « message »), radio *útvarp*, électricité *rafmagn*.

## \* Italien – *italiano*

### Famille

langue romane (famille indo-européenne)

### Localisation

Italie, Suisse, mais aussi USA, Canada, Argentine, Brésil (communautés émigrées)

### Nombre de locuteurs

environ 60 millions

### Système d'écriture

alphabet latin

Il est difficile de déterminer avec précision le nombre de locuteurs de l'italien. Si les 57 millions d'Italiens le comprennent, nombre d'entre eux – dans les campagnes en particulier – parlent des variantes dialectales (*vénitien, pugliese, napolitain, lombard, sicilien, piémontais,...*). Les locuteurs de dialectes différents ont souvent du mal à se comprendre et la langue italienne officielle est fréquemment perçue comme distincte de la langue maternelle. Ainsi, selon certaines sources (SIL), seule la moitié de la population parlerait l'italien standard en tant que langue première.

La forme de l'italien écrit standard tel que nous le connaissons aujourd'hui apparaît au XIV<sup>e</sup> siècle et est fondée sur le texte de la Divine Comédie de Dante et sur les travaux de Pétrarque et Boccaccio. Comme ces célèbres écrivains utilisaient le dialecte de Toscane (en particulier le florentin), l'italien moderne standard est essentiellement toscan. Depuis 1870, le dialecte de Rome a considérablement gagné en prestige, mais il n'égale pas le florentin standard.

L'italien est la langue romane (officielle) la plus proche du français. Quelques différences notables sont toutefois à relever:

- l'italien, au contraire du français, est une langue dans laquelle l'accent tonique – habituellement placé sur l'avant-dernière syllabe – peut servir à distinguer des mots ou des sens différents: *impáro* (j'apprends, accent habituel sur l'avant-dernière syllabe) - *imparó* (il apprit, accent sur la dernière syllabe);
- l'adéquation entre la prononciation et l'écriture y est bien plus grande qu'en français: presque toutes les lettres écrites sont prononcées;
- l'italien fait un grand usage de mots « télescopés », « amalgamés »: *così* (comme ceci); *sul* (sur le);
- l'italien n'a pas de voyelles nasales;
- l'italien, comme l'espagnol et contrairement au français, suffixe généralement le pronom objet au verbe: *trovarmi* (me trouver); etc.

## Japonais

## 日本語

### Famille

langue isolée

### Localisation

Japon

### Nombre de locuteurs

environ 125 millions

### Système d'écriture

le japonais utilise 4 types d'écriture selon les cas (*cf. infra*).

Le Japon est un pays linguistiquement très homogène: plus de 99% de la population parle la langue nationale, qui compte cependant différents dialectes régionaux, dont celui d'Okinawa est le plus différent. La langue standard, dénommée *hyoojungo*, correspond au dialecte de Tokyo et a été élaborée au XIX<sup>e</sup> siècle. Il en est résulté une normalisation progressive et une certaine dénégation des variantes régionales.

L'origine de la langue japonaise est mal connue. On en fait le plus souvent une famille originale – les « langues japonaises » – contenant



une douzaine de langues toutes parlées au Japon et qui, excepté le japonais, sont en voie d'extinction. Certains linguistes rattachent cependant le japonais au coréen, voire aux langues altaïques (ou turco-mongoles) avec lesquelles il partage le caractère agglutinant, ou encore avec certaines langues malaises du Sud-Est asiatique dans lesquelles on retrouve certaines racines lexicales et diverses traditions.

Mais le japonais a surtout été influencé par le chinois auquel il a emprunté plus de la moitié de son vocabulaire et, au IV<sup>e</sup> siècle, son système d'écriture idéographique. En fait, le système actuel d'écriture japonais est très complexe. Aux *kanji* initiaux du chinois sont venus s'ajouter, dès le IX<sup>e</sup> siècle, des *kana*, c'est-à-dire des signes qui constituent un alphabet syllabique et qui ont été introduits pour transcrire certaines particularités de la langue japonaise, en particulier les suffixes verbaux et certains sons. De plus, il existe deux systèmes de *kanas* : les *hiragana*, d'abord utilisés par les femmes qui n'avaient pas accès à l'enseignement des *kanji*, et les *katakana* utilisés pour les chiffres, les mots étrangers, les onomatopées...

Le système actuel mêle ainsi les *kanji* – environ 2000 pour les signes courants (dont 1000 que les élèves doivent s'approprier à l'école primaire!), 6000 pour certains ouvrages spécialisés (contre environ 60000 en Chine) –, les *hiragana* (de forme arrondie et au nombre de 46) et les *katakana* (de forme angulaire, également au nombre de 46). En outre, le japonais peut s'écrire de droite à gauche et de haut en bas, comme à l'origine, ou plus rarement horizontalement et de gauche à droite. Enfin, à tout cela s'ajoute encore les caractères *romaji* utilisés pour transcrire les mots japonais en alphabet latin, pour des sigles, plus rarement sur des enseignes et pour classer les mots dans les dictionnaires. Malgré la complexité de ces différents systèmes et la pression actuelle de l'alphabet latin via l'influence de l'anglais et les nouvelles technologies, les Japonais restent toutefois très attachés à leur écriture traditionnelle qui fait partie de leur identité et constitue d'ailleurs un véritable art calligraphique.

En japonais, l'ordre des mots est très différent du français (verbe à la fin, par exemple); les noms sont en général invariables et les verbes ne se conjuguent qu'en temps et en mode (pas en personne). Le japonais se distingue également par un système de politesse particulièrement riche: ainsi, à côté des formes verbales neutres, par exemple *kaku* (écrire), on a une forme polie *kakimasu* avec le suffixe *-masu* de politesse. Les formulations, en conséquence, peuvent varier fortement selon les interlocuteurs en présence (hommes ou femmes, jeunes ou anciens...). La phonétique, en revanche, est simple et ne pose guère de problème de transcription.

## Kurde – *kirmanci*

### Famille

langue iranienne (famille indo-européenne)

### Localisation

Iran, Irak, Syrie, Turquie

### Nombre de locuteurs

environ 11 millions

### Système d'écriture

en Syrie, Irak et Iran, le kurde s'écrit dans une variante de l'alphabet arabe; en Turquie en alphabet latin

Le kurde fait partie de la famille des langues indo-européennes, plus particulièrement de la famille des langues iraniennes. Dans cette dernière famille, on trouve aussi, outre le *persan* (ou *farsi*):

- l'*ossète*, parlé dans le Caucase central,
- le *baloutche*, parlé dans le sud-est iranien et au Pakistan,
- le *pashto* ou *poshtou*, parlé en Afghanistan et au Pakistan,
- le *dari* et le *tadjik*, variétés dialectales du persan parlées respectivement en Afghanistan et au Tadjikistan, etc.

La parenté entre ces langues iraniennes peut être comparée à celle des langues latines entre elles.

On distingue deux grandes variantes de kurde: le *kurmanja* (ou *kurmanji*) parlé par environ 7 millions de personnes (principalement en Turquie) et le *sorani* (ou kurdi) parlé par environ 6 millions de locuteurs (principalement en Irak et Iran).

Le kurde a beaucoup de peine à se faire reconnaître. Il est langue de minorité, en effet, dans tous les pays où il est présent. En Turquie par exemple, le kurde a longtemps été interdit, voire nié, et les locuteurs fortement réprimés au nom d'une idéologie nationale unitaire. Aujourd'hui, toutefois, la situation semble évoluer plus favorablement en raison entre autres de la volonté de ce pays d'entrer dans la Communauté européenne.

## Langue des Signes Française (LSF)

### Famille

langage visuel

### Localisation

Belgique, France, Suisse romande

### Nombre de locuteurs

50 à 100000

### Système d'écriture

alphabet latin

La langue des signes n'a pas été inventée un beau jour par quelqu'un. Comme les langues orales, elle s'est développée avec le temps et continue son évolution. Pour communiquer, les personnes sourdes rendent visible ce qu'elles veulent dire en faisant correspondre un geste à une unité de signification lexicale ou grammaticale (cf. *Paroles en actes*, volume II). Pour certains usages (épeler des noms propres ou des mots nouveaux), elles utilisent en outre un alphabet dans lequel chaque lettre est exécutée par un geste de la main.

La langue des signes française (LSF) est la langue des sourds de Suisse romande, de France et de Belgique francophone – ce qui signifie que la langue des signes n'est pas universelle: chaque communauté linguistique de sourds possède la sienne. Cependant, la structure grammaticale est pratiquement la même dans le monde entier. De plus, une bonne partie du lexique (signes iconiques) est identique ou proche, les concepts abstraits (signes arbitraires) étant exprimés par des signes plus spécifiquement issus de conventions propres à chaque communauté linguistique. De ce fait, les sourds de pays différents peuvent communiquer beaucoup plus facilement que les entendants, par-dessus les frontières linguistiques.

La langue des signes représente donc la langue naturelle des sourds qui, généralement, sont des personnes privées d'audition depuis leur naissance ou leur petite enfance. Le terme « malentendant », en revanche, est en principe réservé aux personnes qui ont perdu une partie plus ou moins importante de leur faculté auditive, souvent à l'âge adulte. Les malentendants et les « devenus sourds » s'efforcent

avant tout de promouvoir la lecture labiale, une technique qui permet de visualiser le message vocal et, ainsi, de ne pas se retrouver isolé du monde des entendants.

## Lingala–lingala

### Famille

langue bantoue

### Localisation

Congo-Brazzaville, République démocratique du Congo, nord de l'Angola

### Nombre de locuteurs

300 000 comme langue première, environ 7 millions en tant que langue première ou seconde

### Système d'écriture

alphabet latin

Le lingala est l'une des quatre langues «nationales» de la République démocratique du Congo avec le *kikongo*, le *tshiluba* et le *swahili* et l'une des plus de 200 langues recensées dans le pays (le français, et peut-être bientôt l'anglais, étant la langue officielle). C'est la langue dans laquelle communiquent le plus spontanément la plupart des Congolais. Le lingala appartient à la famille des langues bantoues qui réunissent plus de la moitié des 1200 langues africaines (*swahili*, *kirundi*, *kinyarwanda*, *kikongo*, *zoulou*, *xhosa*, *sotho*, *tswana*, etc.). Celles-ci partagent toutes la particularité de posséder des classes nominales, c'est-à-dire un système de préfixes qui classent les noms en fonction de certaines propriétés (cf. *La soleil et le lune*, volume II). Ces langues partagent également un important stock de racines lexicales communes.

## Macédonien

македонски

### Famille

langue slave (famille indo-européenne)

### Localisation

Macédoine

### Nombre de locuteurs

1,5 à 2 millions

### Système d'écriture

alphabet cyrillique

Le macédonien est la langue officielle de la République de Macédoine. Il a su s'imposer comme langue d'enseignement, de communication et comme langue littéraire et de culture. Néanmoins, très proche du bulgare, il est considéré par certains (en particulier les Bulgares!) comme étant plutôt un dialecte du bulgare.

Le macédonien est parlé par 60 à 70% de la population du pays, à côté de nombreuses autres langues liées à différentes minorités ethniques: les Albanais (20 à 30%), les Turcs, les Grecs, les Valaques (qui parlent une variété du roumain: l'*aroumain*), les tsiganes (qui parlent l'*arlija*, une des variétés du *romani*), etc. La reconnaissance de ces minorités est inégale et source d'importantes tensions dans le pays. Après une période de quasi-guerre civile, de fragiles accords de paix ont pu être signés en 2001, qui accordent en particulier un statut de co-officialité à l'albanais, désormais reconnu dans les régions où la population albanophone représente plus de 20%.

## Malgache–malagasy

### Famille

langue malayo-polynésienne

### Localisation

Madagascar

### Nombre de locuteurs

environ 9 millions

### Système d'écriture

alphabet latin

Le malgache a été amené dans l'île par des populations indonésiennes, vers le X<sup>e</sup> siècle. Il appartient à la famille malayo-polynésienne (parfois dénommée austronésienne lorsqu'on y inclut certaines langues ethniques de Formose) qui occupe une aire géographique très vaste, de l'île de Madagascar à l'ouest en passant par une partie du Sud-Est asiatique (en particulier l'Indonésie et les Philippines) pour couvrir presque tout le Pacifique (île de Pâques, Hawaï, Nouvelle-Zélande...). Dans cette famille figurent donc également le *malais*, l'*indonésien* (*bahasa indonesia*), le *tagalog* (*filipino*), le *javanais*, le *balinais*, etc. Elle regroupe environ 265 millions de locuteurs, se plaçant ainsi, en termes de population, derrière la famille indo-européenne et la famille sino-tibétaine.

Ces diverses langues possèdent un stock important de vocabulaire commun: par exemple pour les parties du corps (*l'œil* se dit *mata* en indonésien et en tagalog, *maso* en malgache, *maka* en hawaïen) ou pour les noms de nombres (*huit* se dit *walo* en tagalog, *valo* en malgache, *va'u* à Tahiti, *walu* en hawaïen).

La communauté malgache est composée de 18 ethnies principales, chacune parlant sa propre variété de malgache (*merina*, *betsileo*, *bet-simisaraka*, etc.). C'est le parler des Merinas qui a été choisi comme langue officielle, en raison d'une longue tradition d'écriture remontant à la première moitié du XIX<sup>e</sup> siècle.

Le malgache a autrefois été influencé par le sanscrit, l'arabe et les langues bantoues (*swahili*...); plus récemment, à la suite de la colonisation de l'île (de 1895 à 1960), il a emprunté de nombreux mots au français, souvent en les modifiant (*saribao*, le charbon; *lalimoara*, l'armoire, etc.).

En 1972, un régime d'orientation socialiste imposa constitutionnellement le malgache comme seule langue officielle de la République, cherchant par là à éliminer la domination du français. Pourtant, même si le malgache est la seule langue nationale reconnue, le français est encore utilisé aujourd'hui dans différents domaines de la vie publique (tribunaux, administration, journaux par exemple). A l'école primaire et au premier cycle du secondaire, l'enseignement est entièrement malgachisé. Le français n'est enseigné qu'en tant que langue seconde. Mais il devient langue d'enseignement dès le second cycle du secondaire et dans l'enseignement supérieur.

## Mandarin (voir chinois)

## Marathi

मराठी

### Famille

langue indienne (famille indo-européenne)

### Localisation

Inde, Etat de Maharashtra (Bombay), Goa

**Nombre de locuteurs**

environ 65 à 70 millions.

**Système d'écriture**

alphabet devanagari (utilisé à l'origine pour le sanskrit)

Le marathi est une des 16 langues constitutionnelles de l'Inde. C'est la langue (co-)officielle des Etats de Maharashtra (avec l'hindi) et de Goa (avec le konkani) – ce qui signifie que le parlement, l'administration, l'école, le commerce y fonctionnent dans cette langue, à l'oral et à l'écrit. En revanche, les contacts avec le gouvernement fédéral ont lieu en hindi ou en anglais.

Linguistiquement, le marathi est relativement proche des autres langues indo-européennes du pays. Si l'intercompréhension entre elles n'est pas automatique, elle peut être atteinte assez facilement.

Le vocabulaire, cependant, s'est enrichi de nombreux emprunts aux langues dravidiennes voisines (kannada, télougou), à l'arabe et au persan, ainsi qu'à l'anglais, pour le vocabulaire technique.

Le marathi utilise le même système d'écriture que l'hindi : le devanagari. Il possède une littérature ancienne et prestigieuse, dont le plus illustre exemple est certainement le *Dyaneshwari* (commentaire du *Bhagavad Gita*).

**Malayalam**

മലയാളം

**Famille**

langue dravidienne

**Localisation**

Inde (Kerala)

**Nombre de locuteurs**

environ 35 millions

**Système d'écriture**

alphabet malayalam (dérivé de l'alphabet pali)

Le malayalam est l'une des 16 langues constitutionnelles de l'Inde et la langue officielle de l'Etat du Kerala, province dont le taux d'alphabétisation et d'instruction est particulièrement élevé. Il est parlé par la grande majorité de la population de l'Etat, quelle que soit sa religion.

Le malayalam appartient à la famille des langues dravidiennes – avec le *tamoul*, le *télougou*, le *kannada* et d'autres langues de l'Inde – qui compte environ 200 millions de locuteurs. Ces langues permettent une certaine intercompréhension moyennant une initiation ; par exemple, 4 se dit *nalu* en malayalam, *nangu* en tamoul, *nalgu* en kannada et *nalugu* en télougou. Le terme *malayalam* est formé à partir des racines d'origine dravidienne *mala* (= montagne) et *alam* (= côte).

Comme les autres langues dravidiennes, le malayalam comporte de nombreux suffixes pour marquer le pluriel, le genre (masculin, féminin et neutre), etc. Il comporte un système de déclinaison à sept cas, également marqués par des suffixes. Le vocabulaire contient de nombreux emprunts au sanskrit, considérés autrefois comme un signe de raffinement, et à l'anglais, dans le domaine technique en particulier.

Le malayalam possède son propre système d'écriture, issu du système *pali* utilisé autrefois pour les textes sacrés bouddhiques.

**Nahuatl – mexìcatlàtōlli****Famille**

langue uto-aztèque

**Localisation**

Mexique, Salvador

**Nombre de locuteurs**

environ 1 million et demi

**Système d'écriture**

aujourd'hui alphabet latin ; mais, à l'époque précolombienne, on trouvait déjà en développement un système d'écriture idéographique avec certains éléments purement phonétiques.

Le nahuatl est plus connu sous le nom d'*aztèque*, puisqu'il était la langue « officielle » de l'empire aztèque jusqu'à l'arrivée des Espagnols. Il a joué un rôle de langue de contact dans toute l'Amérique centrale jusqu'après la conquête. Son caractère de langue de prestige lui a valu l'attention de plusieurs religieux grammairiens (la première grammaire date de 1547) et a donné lieu à une importante littérature entre 1550 et 1650 (nahuatl dit « classique »).

Il est aujourd'hui parlé, sous diverses formes dialectales, dans plusieurs régions du Mexique (District Federal, Morelos, Guerrero, Veracruz, Hidalgo, Michoacan, Jalisco) et dans la République du Salvador. En revanche, il a largement perdu son prestige et n'est guère reconnu ni enseigné malgré l'existence de lois qui devraient en assurer la sauvegarde.

Linguistiquement, on peut signaler la formation de certains pluriels par redoublement de la syllabe initiale (*coyotl* (= coyote), pl. *cocoyo*) et l'usage d'un système de numération quinaire (base 5) et vicésimal (base 20). Quelques mots français proviennent du nahuatl, via l'espagnol : *coyote*, *ocelot*, *chocolat*, *tomate*, *avocat*.

**Néerlandais – nederlands****Famille**

langue germanique (famille indo-européenne)

**Localisation**

Pays-Bas, Belgique, Surinam, Caraïbes et Antilles hollandaises

**Nombre de locuteurs**

environ 20 millions

**Système d'écriture**

alphabet latin

Le néerlandais est la langue des Pays-Bas (la Hollande) et, à parité avec le français, l'une des deux langues de la Belgique. Son nom local est *nederlands*, mais l'appellation de *flamand* (*vlaams*) est utilisée en Belgique. Les différences entre néerlandais et flamand sont minimes, d'autant plus que, depuis 1935, une commission mixte belgo-néerlandaise travaille à une standardisation de l'orthographe. L'*afrikaans*, langue officielle de l'Afrique du Sud apportée par les colons, est à l'origine du néerlandais, mais qui a évolué différemment et s'est simplifié au contact d'autres populations. L'intercompréhension entre néerlandais et afrikaans reste possible mais non sans attention.

Situé entre l'allemand et l'anglais, le néerlandais est plus proche du premier. La grammaire, en particulier, s'en rapproche fortement (subordonnées avec inversion, particules verbales séparables...).

Grâce à la longue tradition maritime des Hollandais, le néerlandais a laissé son empreinte sur de nombreuses langues, en particulier dans le domaine nautique (en français par exemple, les mots *bâbord*, *matelot*).

**Népal**

नेपाली

**Famille**

langue indienne (famille indo-européenne)

**Localisation**

Népal, Etats indiens (ouest du Bengale, Assam et Sikkim) et sud du Bouthan

**Nombre de locuteurs**

environ 16 millions

**Système d'écriture**

alphabet devanagari (utilisé à l'origine pour le sanskrit)

Le népal ou népalais est la langue officielle du Népal. Il n'y est toutefois parlé que par moins de la moitié de la population totale, cohabitant avec de nombreuses autres langues indo-européennes (*awadhi*, *bhojpuri*, *hindi*, *maithili*, *tharu*) ou non (sino-tibétaines: *gurung*, *limbu*, *magar*, *newari*, *tamang*, *tibétain*; austro-asiatiques: *mundari*, *santali*; dravidiennes: *kurux*). Le népal regroupe en outre plusieurs dialectes qui peuvent s'éloigner assez fortement de la langue standard. La grammaire du népal fonctionne tout à fait comme celle de l'hindi et l'alphabet est le même.

**Norvégien – bokmål / riksmål (norsk)****Famille**

langue germanique (branche scandinave) de la famille indo-européenne

**Localisation**

Norvège

**Nombre de locuteurs**

environ 5 millions

**Système d'écriture**

alphabet latin (complété par quelques graphies originales: â = [ʃ]; ø = [ø]...)

Outre quelques langues minoritaires (*sami*, *finnois*), la Norvège se caractérise par la cohabitation de nombreux dialectes qui ont fait l'objet, depuis le XIX<sup>e</sup> siècle, d'une politique linguistique constante. Deux variétés ont d'abord été standardisées: le *riksmål* («langue du royaume») parlé principalement dans les villes du sud, fortement influencé par le danois, et le *landsmål* («langue du pays», «langue de la campagne»), sorte de synthèse des différents dialectes parlés dans le pays. Dès 1885, ces deux variétés – entre lesquelles la compréhension est relativement aisée – étaient devenues langues officielles du pays. Après une tentative échouée d'unification dans un «norvégien commun» (*sammorsk*), les deux variétés aujourd'hui connues sous le nom de *bokmål* («langue du livre», parlée par plus de 70% de la population) et de *nynorsk* («nouveau norvégien», parlé par un peu moins de 30% de la population) sont encore en usage, toutes deux enseignées dans les écoles même si la première tend à dominer. Entre autres effets, cette attention constante portée aux deux variétés a conduit à les adapter et les réformer régulièrement, par exemple au niveau du lexique et de l'orthographe.

En tant que langues scandinaves, les deux variétés du norvégien restent proches des autres langues de ce groupe, du danois en particulier pour le *bokmål*, du janska (dialecte parlé en Suède) pour le *nynorsk*.

**\* Ourdou (urdu)**

اردو

**Famille**

langue indienne (famille indo-européenne) très proche de l'hindi

**Localisation**

Pakistan et Inde (Ile Maurice, Afrique du Sud)

**Nombre de locuteurs**

environ 60 millions

**Système d'écriture**

alphabet arabe adapté («Nastaliq»), s'écrit de droite à gauche

Malgré leurs noms différents, l'hindi et l'ourdou constituent deux variétés très proches l'une de l'autre, car ils ont longtemps constitué la même langue (celle qui était en usage à la cour des souverains moghols, au XVIII<sup>e</sup> siècle). Aujourd'hui, le vocabulaire hindi tend à dériver principalement du sanskrit (en raison d'une volonté sociolinguistique de «désislamisation»), tandis que l'ourdou contient de nombreux mots d'origine persane et arabe. De plus, l'hindi utilise l'alphabet devanagari alors que l'ourdou privilégie un alphabet arabe modifié d'influence persane. Enfin, en Inde comme au Pakistan, l'hindi est parlé principalement par les hindous, tandis que l'ourdou est utilisé essentiellement par les musulmans. Ces questions de délimitation et de dénomination ont fait l'objet de débats afin de savoir si on devait les considérer comme une seule langue ou non.

Langue nationale du Pakistan, langue de l'enseignement public, l'ourdou n'y est en fait parlé en tant que langue maternelle que par un peu plus de 10 millions de locuteurs, autrement dit à peu près 10% de la population (la majorité de la population parle plutôt pandjabi, une autre langue indo-européenne). En revanche, il y est utilisé comme langue seconde par une part importante de la population et constitue également la langue de l'appartenance religieuse. En Inde, parlé par environ 50 millions de locuteurs, il a le statut de langue «constitutionnelle» et il fonctionne comme langue officielle ou co-officielle dans plusieurs Etats: Uttar Pradesh, Bihar, Andhra Pradesh, Kashmir.

**Pandjabi – punjabi**

ਪੰਜਾਬੀ

**Famille**

langue indienne (famille indo-européenne)

**Localisation**

Inde (Pandjab), Pakistan

**Nombre de locuteurs**

environ 27 millions (80 millions si l'on compte les locuteurs pakistanais)

**Système d'écriture**

alphabet gurumukhi (variante du devanagari)

Parlé en Inde et au Pakistan, le pandjabi se trouve au cœur d'un violent conflit politique! Le pandjabi, en effet, est l'une des langues constitutionnelles de l'Inde. Mais il est également la langue maternelle d'environ 60% de la population du Pakistan, loin devant l'ourdou. En fait, la situation linguistique du Pakistan est complexe car pandjabi et ourdou (et donc également hindi) sont très proches, mais ils s'écrivent dans des systèmes différents, gurumukhi pour le premier, adapté de l'alphabet arabe pour le second. Or, seul l'ourdou est enseigné au Pakistan et les locuteurs pakistanais de pandjabi écrivent essentiellement en ourdou. C'est pourquoi il est difficile d'estimer précisément le nombre de locuteurs de pandjabi.



En Inde, le pandjabi est parlé dans l'Etat du Pandjab où il est entre autres la langue des Sikhs. Là il s'écrit selon l'alphabet gurumukhi, un système dérivé du système devanagari (utilisé en hindi) dont il reste toutefois suffisamment proche pour qu'on puisse l'acquérir assez facilement. La grammaire et la phonétique du pandjabi ne se distinguent guère non plus de celles de l'hindi/ourdou.

## \* Persan (farsi)

فارسی

### Famille

langue iranienne (famille indo-européenne)

### Localisation

Iran, Afghanistan, Tadjikistan, Pakistan (Irak, Qatar, Emirats arabes unis, Oman)

### Nombre de locuteurs

environ 30 millions

### Système d'écriture

alphabet arabe légèrement adapté

Le persan, également appelé *farsi*, est une langue de la famille iranienne, qui contient également le *kurde*, l'*ossète*, le *baloutche*, le *pashto*, etc. La parenté de ces langues entre elles peut être comparée à celle qu'on observe entre les langues romanes. Le *dari* parlé par 6 millions d'Afghans et 1 million de Pakistanais et le *tadjik* parlé par 3 millions de locuteurs peuvent être considérés comme des variétés dialectales du persan. Ils présentent toutefois des différences phonétiques et lexicales qui peuvent causer quelques difficultés dans l'intercompréhension. En outre, le tadjik s'écrit en alphabet cyrillique. En revanche, la langue littéraire est sensiblement la même. Le terme *farsi* (du même nom que la province de Fars en Iran) est le même que Perse (car [f] et [p] sont des sons proches) et évoque les Parthes de l'Antiquité.

La langue farsi est la langue nationale de l'Iran, enseignée à l'école, utilisée dans les médias et la littérature; elle y est parlée par plus de 20 millions de locuteurs (un peu moins de 40% de la population). Elle s'écrit dans un alphabet arabe adapté, afin de pouvoir représenter quelques sons uniquement présents dans le farsi ([p], [tʃ], [ʒ]...).

Phonétiquement, le farsi est une langue indo-européenne et ne présente pas de grandes particularités, si ce n'est le [q] arabe, sorte de [k] prononcé très en arrière de la gorge; la longueur des voyelles possède en outre un caractère distinctif. Grammaticalement aussi, il est clairement une langue indo-européenne. Lexicalement, il est toutefois très intéressant car, situé au milieu de l'aire indo-européenne, il constitue en quelque sorte un pont entre les langues indo-européennes de l'ouest et celles de l'est. Le farsi comporte en outre un système complexe de formes alternatives pour exprimer les nuances des relations hiérarchiques et de politesse.

Le français a emprunté de nombreux mots au persan (*divan*, *douane*, *caviar*, *habouche*, *échec*, *mat...*) et lui en a également fourni beaucoup.

## Polonais–polski

### Famille

langue slave (famille indo-européenne)

### Localisation

Pologne, Lituanie, Ukraine

### Nombre de locuteurs

environ 40 millions

### Système d'écriture

alphabet latin (possède de nombreux signes diacritiques ajoutés aux lettres latines pour en modifier la réalisation)

Le polonais, langue officielle de la Pologne, est également parlé par une importante diaspora dans le monde (minorités en Lituanie, en Russie, en Biélorussie...). C'est une langue slave proche du biélorusse et du slovaque, qui s'écrit avec un alphabet latin adapté au système phonétique du polonais; il comporte entre autres des voyelles nasalisées qui sont indiquées par une cédille placée sous les lettres a et e, des consonnes doubles qui ne correspondent qu'à un seul son (ch [x], cz [tʃ], rz [ʒ]...). D'apparence très difficile à décrypter en raison du nombre important de consonnes que semblent comporter les mots, la graphie du polonais est en fait régulière et assez fidèle à la prononciation.

Phonétiquement, le polonais possède quelques sons caractéristiques (entre autres un l palatalisé noté *ł*, qui se prononce un peu comme le son [u]). Grammaticalement, il partage la plupart des caractéristiques des autres langues slaves (système de sept cas avec déclinaisons, variation des formes verbales en fonction de l'aspect perfectif ou imperfectif, etc.). Le vocabulaire est également proche des autres langues slaves. De nombreux emprunts proviennent de l'allemand et du français (*rachunek* → Rechnung, *szofer* → chauffeur, etc.).

## \* Portugais–português

### Famille

langue romane (famille indo-européenne)

### Localisation

Angola, Brésil, Guinée-Bissau, îles du Cap-Vert, Mozambique, Portugal, république de São Tomé et Príncipe

### Nombre de locuteurs

environ 176 millions (dont 10 millions au Portugal)

### Système d'écriture

alphabet latin

Le portugais est la cinquième langue la plus parlée dans le monde (en tant que langue première). En dehors du Portugal, il est parlé dans cinq pays d'Afrique et au Brésil (anciennes colonies).

On observe entre le portugais du Brésil (*português do Brasil*) et le portugais du Portugal (*português de Portugal*) des différences qui sont comparables aux différences existant entre le français de France et celui du Québec ou entre l'anglais britannique et l'anglais américain. Il existait jusqu'à récemment deux codes écrits relativement distincts, mais les deux pays ont signé des accords et unifié leur orthographe. Les langues *parlées* du Brésil et du Portugal se distinguent plus nettement; la prononciation brésilienne, par exemple, est moins « chuintante » (sons [j], [ʒ]), moins nasalisée (voyelles). Après l'indépendance du Brésil, les emprunts aux langues amérindiennes et africaines (principalement le *yorouba*, langue la plus fréquemment parlée par les anciens esclaves au Brésil) ainsi que les créations brésiliennes sont devenus très courants au Brésil.

Par ailleurs, le Portugal a connu de grandes vagues d'émigration. Ainsi, selon le recensement fédéral de 1990, la Suisse compte à peine moins de 100 000 lusophones, migrants pour la plupart.

## Rom (langues tsiganes)–romani

### Famille

langue indienne de la famille indo-européenne

**Localisation**

de l'Asie à l'Europe et aux Amériques

**Nombre de locuteurs**

environ 1,5 million

**Système d'écriture**

alphabet latin

La langue *romani* est l'unique représentante européenne des langues indiennes appartenant à la famille indo-européenne. Elle a préservé en grande partie l'héritage des langues de l'Inde du Nord, plus particulièrement le hindi et le rajasthani avec lesquels elle possède en commun environ 60% du vocabulaire de base.

La langue *romani* est parlée par diverses populations d'origine indienne, la plupart nomades, qu'on dénomme Tsiganes, mais aussi Romani, Roms, Bohémiens, Gitans, Manouches, etc. Venus du nord de l'Inde aux IX<sup>e</sup> et XIII<sup>e</sup> siècles, elles ont formé la *Romani Cel* (d'où leur surnom de «Romanichels»), mais ils se nomment eux-mêmes *Romané Chavé* «fils de Ram» (héros de l'épopée indienne Ramanaya).

Les Tsiganes n'ont par conséquent pas d'Etat propre, ils sont dispersés non seulement à travers l'Europe mais aussi en Amérique (Argentine, Brésil, Colombie, Etats-Unis) et en Asie. Bien qu'on ne possède pas de recensement précis, ni pour la langue ni pour la population, on peut estimer leur nombre à environ 80 millions, dont la plupart toutefois ne parlent plus la langue ancestrale.

Il existe de plus de nombreuses variétés de romani, influencées par les langues des pays dans lesquels elles sont parlées, et entre lesquelles la compréhension n'est pas toujours possible. En voici quelques exemples: *romani* ou *arlija* (Europe de l'Est; 500 000 locuteurs), *vłax romani* (Europe de l'Est, mais aussi Colombie; 300 000), *romani des Carpathes* (Carpathes; 400 000), *sinte* (Europe centrale, dont la Suisse; 100 000), etc. En tout, on peut estimer le nombre de locuteurs à 1 million et demi, quoique Malherbe (1985, 1603) parle de 10 millions... En Espagne, au Portugal et au Brésil, on trouve des Tsiganes parlant le *caló* qui est une variante dialectale basée sur les langues romanes et comprenant moins de 300 mots d'origine romani.

**\* Romanche – rumantsch grischun****Famille**

langue romane formant avec le ladin et le frioulan le sous-groupe des langues rhéto-romanes (famille indo-européenne)

**Localisation**

Suisse, canton des Grisons

**Nombre de locuteurs**

environ 40 000 personnes qui mentionnent le romanche comme langue «principale» (celle qu'ils savent le mieux), 66 000 si l'on compte ceux qui disent l'utiliser régulièrement dans certains contextes (cf. données du recensement fédéral 1990)

**Système d'écriture**

alphabet latin

Le romanche est, depuis 1938, une des quatre langues nationales de la Suisse. Selon la nouvelle Constitution, il a également le statut de langue semi-officielle, c'est-à-dire qu'il est «langue officielle pour les rapports que la Confédération entretient avec les personnes de langue romanche» (cf. article 70 de la nouvelle Constitution).

Le romanche comporte cinq variantes régionales, toutes parlées et écrites: le *sursilvan*, dans la vallée du Rhin antérieur (Disentis), le *sutsilvan* et le *surmiran* dans le centre du canton, enfin le *putèr* et le *vallader* qui forment dans l'Engadine et dans le Val Müstair le sud-est de l'aire romanche et qui sont proches du ladin parlé en Italie. Le romanche se maintient dans des proportions variables selon les régions, davantage en Surselva, en Basse-Engadine, dans le Val Müstair et dans la partie sud-est des Grisons centraux où il conserve en général la majorité (jusqu'à plus de 90%) face à l'allemand, moins dans la Haute-Engadine et dans la Sutselva où, sous la pression de l'allemand et en raison entre autres du tourisme, son recul est particulièrement sensible.

Globalement, on peut toutefois parler d'une menace réelle pesant sur la quatrième langue nationale. C'est pourquoi diverses initiatives et mesures ont été prises au niveau cantonal et fédéral. Dans le canton, la «Lia Rumantscha» a créé un romanche normalisé, le *rumantsch grischun*, basé essentiellement sur trois des cinq variétés: le vallader, le sursilvan et le putèr. Cette langue standardisée et unifiée, mais parfois perçue comme artificielle, est utilisée – surtout à l'écrit – dans différents domaines jusque-là exclusivement réservés à l'allemand et à l'italien: l'administration, l'affichage, la signalisation, les médias et, à un degré moindre, les écoles. La Confédération apporte son soutien à ces initiatives, comme le lui enjoint l'article 70 de la nouvelle Constitution, afin de «sauvegarder et promouvoir le romanche».

Cf. présentation plus détaillée dans les annexes des activités *P'tit déj*, volume I, et *Schi vain ün auto I*, volume II.

**Roumain – română****Famille**

langue romane (famille indo-européenne)

**Localisation**

Roumanie, Moldavie, certaines régions de Serbie, du Monténégro, de Croatie, de Slovénie, de Bosnie-Herzégovine, de Bulgarie et de Hongrie

**Nombre de locuteurs**

environ 26 millions

**Système d'écriture**

alphabet latin (jusqu'en 1860, le roumain s'écrivait en alphabet cyrillique)

Le roumain, la plus orientale des langues romanes, est le fruit d'une colonisation et d'une romanisation qui eut lieu au début du II<sup>e</sup> siècle. Ainsi qu'en témoignent de nombreuses traces dans la langue, les légionnaires semblent être venus surtout de Sicile et de Sardaigne.

Mais, au cours de son histoire, cette langue a subi de nombreuses autres influences – de langues non romanes, et même non indo-européennes – qui ont profondément marqué son vocabulaire et, parfois, sa structure. Le roumain partage ainsi certains traits avec d'autres langues balkaniques (grec, albanais, langues slaves), avec le hongrois, avec le turc, avec l'allemand via les populations germaniques implantées en Transylvanie. D'ailleurs le roumain s'est écrit d'abord avec l'alphabet cyrillique, jusqu'en 1860, et ce n'est qu'en 1990 que la Moldavie a opté pour l'alphabet latin. En outre, le roumain conserve quelques traces de l'ancienne langue dace (vraisemblablement apparentée à l'albanais) parlée avant l'invasion romaine.



Parmi ces traits particuliers, qui concernent avant tout le lexique (oui se dit *da*, comme en russe!), on signalera la suffixation de l'article après le nom, la fréquence d'emploi du subjonctif, la formation du futur avec un auxiliaire «vouloir», etc. Le roumain est par ailleurs la seule langue latine à avoir conservé des cas et des déclinaisons.

## \* Russe

РУССКИЙ

### Famille

langue slave (famille indo-européenne)

### Localisation

Fédération de Russie, anciens pays de l'Union soviétique, Israël, Etats-Unis

### Nombre de locuteurs

environ 165 millions

### Système d'écriture

alphabet cyrillique

Le russe (*russky*) est bien sûr la langue numériquement la plus importante des langues slaves, entre lesquelles les ressemblances sont généralement très étroites. Il est langue maternelle d'environ 165 millions de personnes, en Russie, mais aussi en Israël où s'est installée une importante communauté russophone et dans les anciennes républiques de l'URSS. Le russe est également langue seconde et véhiculaire dans l'ensemble des pays de l'ancien bloc de l'Est, même s'il est fortement menacé dans tous ces pays par les langues occidentales, l'anglais en premier lieu.

On soulignera que la Russie inclut un nombre important d'autres langues : plus de 100, qui jouissent pour la plupart d'une reconnaissance officielle ; certaines d'entre elles sont également des langues slaves (ukrainien, biélorusse...), d'autres sont des langues indo-européennes (langues baltées, iraniennes...) et d'autres appartiennent à d'autres familles (turques, finno-ougriennes, mongoles) ou sont isolées (tchétchène, ingouche, lak, eskimo...).

Voici quelques-unes des principales caractéristiques phonétiques et grammaticales de la langue russe : l'existence d'un [r] roulé très perceptible, le nombre important de sons «chuintants» ([ʃ], [tʃ], [tʃʃ]) l'existence d'un accent tonique distinctif (mais qui n'est pas indiqué dans la graphie), l'absence d'article, l'existence d'un genre neutre, un système de déclinaison à six cas et un système verbal assez complexe exprimant des nuances auxquelles nous ne sommes guère habitués, en particulier au niveau aspectuel (perfectif/imperfectif...).

Le vocabulaire est très riche et contient de nombreux emprunts à l'allemand, au turc et, surtout, au français qui a exercé une influence considérable sur cette langue, à l'époque de Catherine II en particulier.

L'alphabet cyrillique, utilisé pour la langue russe, a été créé au IX<sup>e</sup> siècle, à partir de l'alphabet grec, par les apôtres Cyrille et Méthode pour les besoins de l'évangélisation.

## \*Serbo-croate

(ou : quelques données sur la langue serbe, croate, bosniaque et monténégrine)

### Famille

langues slaves du Sud (famille indo-européenne)

### Localisation

Serbie et Monténégro, Croatie, Bosnie et Herzégovine

### Nombre de locuteurs

environ 16 millions

### Système d'écriture

alphabets latin et cyrillique

Du temps de la République socialiste fédérative de Yougoslavie (1943-1992), le serbo-croate ou croato-serbe était la langue maternelle de 16 millions d'habitants de quatre des Républiques de ce pays (la Serbie, la Croatie, la Bosnie-Herzégovine et le Monténégro), ainsi que d'un certain nombre d'habitants des autres régions du pays (Républiques de Macédoine et Slovénie), de minorités dans les pays avoisinants et parmi la diaspora éparpillée de par le monde. A cette langue correspondaient deux écritures : latine en Croatie et en Bosnie-Herzégovine, cyrillique en Serbie et au Monténégro. Après le démembrement de la RSFY (1992) et dans le but d'affirmer leurs identités nationales, les pays nouvellement indépendants issus des anciennes Républiques yougoslaves ont repris et renforcé leurs variantes spécifiques par rapport à l'ancienne langue commune. Ainsi, l'actuelle langue officielle en Serbie est le serbe, en Croatie le croate, en Bosnie et Herzégovine (Etat lui-même formé de deux «entités» : la Fédération de Bosnie et Herzégovine, avec pour capitale Sarajevo, et la République serbe dont Banja Luka est la capitale) le bosniaque, le serbe et le croate. Au Monténégro, on continue officiellement de parler le serbo-croate, mais les demandes pour que la variante monténégrine soit reconnue comme une langue à part entière se font de plus en plus pressantes. En ce qui concerne l'écriture, le cyrillique a complètement été rejeté de Croatie (les documents en cyrillique doivent être «traduits» en écriture latine par un notaire) et, dans une moindre mesure, de la Fédération de Bosnie et Herzégovine (l'une des deux entités de Bosnie et Herzégovine). En Serbie et plus encore en République serbe (l'autre entité de Bosnie et Herzégovine), l'utilisation publique et administrative du cyrillique est encouragée, cependant l'écriture latine reste également en usage et des livres et journaux, par exemple, sont imprimés en écriture latine.

Lorsqu'on parle de cette langue ou de ces langues, il importe donc de distinguer entre la situation linguistique d'une part et politico-linguistique de l'autre.

D'un point de vue strictement linguistique, les langues serbe, croate et bosniaque sont toutes des langues slaves du Sud et répondent pratiquement aux mêmes règles morphologiques (flexion, dérivation), syntaxiques et prosodiques. Les différences existent, mais elles sont négligeables en comparaison de la masse des similitudes. Les différences lexicales peuvent être assimilées à celles existant entre le français du Québec et celui de France ou encore entre l'anglais britannique et américain. En tout état de cause, les locuteurs de la langue standard littéraire croate et ceux de la langue standard littéraire serbe se comprennent mutuellement mieux que ne se comprennent les locuteurs de certains dialectes à l'intérieur même de la langue croate ou, respectivement, de la langue serbe.

## Slovaque – slovenský

### Famille

langue slave (famille indo-européenne)

### Localisation

Slovaquie

### Nombre de locuteurs

environ 5 millions

**Système d'écriture**

alphabet latin

Le slovaque est la langue nationale de la Slovaquie. Il est très proche du tchèque, ainsi que dans une mesure légèrement moindre du polonais et de l'ukrainien. L'intercompréhension entre ces langues ne pose guère de problèmes. Le slovaque n'a été véritablement reconnu que depuis l'avènement de la Tchécoslovaquie en 1918. Avant, malgré l'existence d'une littérature et un usage largement partagé par toutes les couches de la population, il était surtout réservé à l'usage privé.

Le slovaque utilise l'alphabet latin complété de quelques signes qu'on retrouve, avec quelques différences, dans d'autres langues slaves (tchèque, croate, slovène): ě (= [tʃ]), š (= [ʃ]), ž (= [ʒ])...

**Slovène – slovenia****Famille**

langue slave (famille indo-européenne)

**Localisation**

Slovénie, Autriche (Carinthie), Italie

**Nombre de locuteurs**

environ 2 millions

**Système d'écriture**

alphabet latin

Le slovène est la langue officielle de la république de Slovénie. Historiquement, les Slovènes ont été isolés des autres peuples slaves du Sud et leur langue a conservé certains caractères archaïques, en particulier l'existence, entre le singulier et le pluriel, d'un nombre « duel », désignant deux personnes ou deux choses et marqué dans les noms, les adjectifs, les pronoms et la conjugaison verbale.

A l'époque austro-hongroise déjà, les Slovènes ne manquaient pas d'élites lettrées et le clergé a joué un rôle important dans l'élaboration progressive d'une norme littéraire de la langue. La première grammaire slovène date du XIX<sup>e</sup> siècle, publiée sous un titre qui délimitait le territoire linguistique du slovène, tout en reconnaissant la diversité des pratiques dialectales. Il fallut néanmoins attendre la reconnaissance d'un peuple slovène et la création au sein de la Yougoslavie d'une république fédérée de Slovénie (1945) pour qu'une langue slovène unifiée s'impose véritablement, non plus seulement comme norme théorique et savante, mais comme outil généralisé de communication. Comme les autres langues slaves des régions catholiques, le slovène utilise l'alphabet latin complété par quelques signes particuliers (č, š, ž ... ; cf. slovaque).

Le slovène, qui compte différents dialectes relativement distincts les uns des autres, est parlé par près de 90% des Slovènes en tant que langue maternelle. Il est également parlé par une minorité installée en Carinthie autrichienne (20 à 40 000 locuteurs; le nom de la capitale, Graz, est d'origine slave et évoque le mot *grad*, ville), dénommée « Windisch », dont la langue est fortement influencée par l'allemand. Enfin, près de 100 000 personnes parlent le slovène en Italie, dans la région du Frioul surtout où cette langue jouit d'un statut officiel et est enseignée, à tous les niveaux, dans certaines écoles.

**Source**

Jean-Arnault Dérens et Catherine Samary (2001). *Les conflits yougoslaves de A à Z*. Paris, Editions de l'atelier & Editions ouvrières, pp. 184-187.

**\* Somali – af-soomaali****Famille**

langue couchitique (famille chamito-sémitique)

**Localisation**

Somalie (Corne de l'Afrique), Ethiopie, Kenya

**Nombre de locuteurs**

environ 10 millions

**Système d'écriture**

alphabet latin

Le somali est parlé en Somalie (plus de 60% de la population), en Ethiopie (env. 1 million) et au Kenya (300 000).

La Somalie fut une colonie anglaise et italienne jusqu'en 1960, année de la création de la République Démocratique de Somalie. Dès 1972, le gouvernement militaire officialise le somali véhiculaire comme seule langue de l'Etat. En 1973, il impose une nouvelle écriture basée sur l'alphabet latin; l'ancien alphabet arabe doit alors faire place à une écriture proprement somalienne, distincte des pays arabophones voisins. Le somali devient ainsi l'une des rares langues chamito-sémitiques à être écrites en alphabet latin et non en alphabet arabe.

La Somalie est également l'un des rares Etats africains à avoir fait table rase des langues coloniales: toute la scolarisation obligatoire se fait uniquement en somali et les langues étrangères sont très peu enseignées.

La situation actuelle de la Somalie est très délicate. Depuis 1992, le pays a été coupé en deux. Le Nord, ancienne colonie anglaise, a proclamé son indépendance sous son nom colonial de *Somaliland*; quant au Sud, il a basculé dans une guerre civile permanente.

**Suédois – svenska****Famille**

langue germanique (branche scandinave) de la famille indo-européenne

**Localisation**

Suède et côte sud-ouest de la Finlande

**Nombre de locuteurs**

environ 9 millions

**Système d'écriture**

alphabet latin

Le suédois est aujourd'hui la plus importante des langues scandinaves qui, de manière générale, sont restées très proches les unes des autres (excepté l'islandais et le féroïen). Ainsi, l'intercompréhension, à l'écrit du moins, est possible avec le danois et le norvégien. Le suédois compte quelques variétés: le *jamska* (très proche du nynorsk de la Norvège) au nord ouest, le *skåne* au sud, le *dalecarlien* (autour de Falun) et la variété parlée en Finlande – où elle est reconnue comme deuxième langue officielle du pays – par plus de 300 000 locuteurs.

Comme les autres langues scandinaves, le suédois connaît un article défini suffixé au nom. Il possède en outre, à côté du neutre, un genre « commun » intégrant le masculin et le féminin, un cas (le génitif) et, surtout, un accent tonal à fonction distinctive, qui porte à la fois sur l'intensité et sur des variations de hauteur mélodique (à l'instar des tons dans des langues comme le chinois).



L'alphabet comporte quelques particularités (â = o long, ö = [ø], sj et sk = [ʃ]...) qui rendent parfois difficile le repérage de mots qu'on connaît pourtant : ainsi *fâtölj* et *byrà*, dans lesquels on retrouve respectivement fauteuil et bureau...

### \* Suisse-allemand – *Schwyzerdütsch*

#### Famille

langue (ou, plutôt, ensemble de dialectes) germanique de la famille indo-européenne

#### Localisation

Suisse alémanique et régions avoisinantes (dans la mesure où il n'y a pas de frontière linguistique stricte entre les dialectes de la Suisse et ceux de l'Alsace, de l'Autriche et des régions allemandes voisines)

#### Nombre de locuteurs

environ 4 à 5 millions (64% de la population suisse)

#### Système d'écriture

alphabet latin

Bien que l'Etat helvétique ait été exclusivement germanophone jusqu'en 1798, le suisse allemand n'a jamais été reconnu officiellement. Mais cela ne l'a pas empêché de perdurer sans problème en tant que langue essentiellement orale (fait suffisamment rare pour être relevé!). Cette pérennité est vraisemblablement due, entre autres raisons, à la tradition germanophone peu centralisatrice et peu encline à imposer une langue unique à toutes les régions.

La situation linguistique de la Suisse alémanique est caractéristique de ce que les linguistes nomment « diglossie » : c'est à dire que l'allemand standard et le suisse-allemand coexistent, mais chacun est utilisé dans des domaines déterminés. Aujourd'hui, le choix de l'un ou l'autre dépend essentiellement du canal de communication : à l'écrit l'allemand et à l'oral le dialecte. Le suisse-allemand peut s'écrire (dans les SMS par exemple) mais, du fait de cette situation diglossique et de l'existence de variétés dialectales, il n'existe pas de norme écrite stabilisée.

Les parlers de la Suisse alémanique sont tous issus du *vieux-haut-allemand*. Mais ils sont bien loin de constituer un parler uniforme et les Suisses allemands seraient bien en peine de dire combien leur région compte de dialectes différents. D'habitude, ils désignent les dialectes selon les cantons, ce qui permet une caractérisation relativement satisfaisante dans la mesure où les frontières cantonales coïncident fréquemment avec une frontière historique, confessionnelle et/ou géographique. La distance avec l'allemand standard n'est pas la même pour tous les dialectes : ceux de l'Est, en particulier, sont plus proches de l'allemand.

Cf. Annexe de l'activité *Schi vain ün auto II*, volume II.

### Swahili – *kiswahili*

#### Famille

langue bantoue

#### Localisation

Afrique de l'Est, entre autre le long des côtes : Tanzanie, Kenya, Ouganda, Zaïre, Comores, Zanzibar

#### Nombre de locuteurs

environ 5 millions (plus de 30 millions comme langue seconde véhiculaire)

#### Système d'écriture

alphabet latin

Le swahili est la langue officielle de la Tanzanie, mais il n'y est parlé comme langue maternelle que par une petite partie de la population. Sa grande importance provient du fait qu'il s'est imposé comme langue véhiculaire permettant la communication entre les locuteurs d'ethnies différentes qui, pour la plupart, parlent également des langues bantoues. Ainsi le swahili peut être compris depuis le Kenya jusqu'à l'Ouganda ou au Congo.

Le swahili appartient à la famille des langues bantoues qui réunissent plus de la moitié des 1200 langues africaines (kirundi, kinyarwanda, kikongo, lingala, zoulou, xhosa, sotho, tswana, etc.). Celles-ci partagent toutes la particularité de posséder des classes nominales, c'est-à-dire un système de préfixes qui classent les noms en fonction de certaines propriétés ; c'est pourquoi, par exemple, tous les noms de langues sont précédés du préfixe *ki-*. Ces langues partagent également un important stock de racines lexicales communes.

Cf. Annexe de l'activité 13, vol. II : *Et pourquoi pas « la » soleil et « le » lune ?*

### \* Tagalog (ou filipino ou pilipino)

#### Famille

langue malayo-polynésienne

#### Localisation

Philippines

#### Nombre de locuteurs

17 millions

#### Système d'écriture

alphabet latin

Langue nationale des Philippines, enseigné dans les écoles, le filipino (nom « officialisé » du tagalog qui était d'abord la langue de l'ethnie tagale, dans la région de Manille) n'est cependant la langue maternelle que d'environ un quart de la population ; en revanche, si l'on inclut également tous les locuteurs qui l'utilisent, à des degrés très divers, comme langue seconde, ce nombre s'élève à plus de 40 millions. Le pays, formé de plus de 7000 îles, compte une septantaine de langues (dont le *visayan*, l'*ilokano*, le *chinois*...). Le filipino a été institué comme langue nationale et dénommé ainsi avec l'idée qu'il devra s'enrichir à partir des autres langues parlées dans le pays. Mais il peine en fait encore à s'imposer et est fortement concurrencé par l'anglais (langue utilisée après la guerre hispano-américaine, en 1898, et la défaite des anciens colonisateurs), qui domine dans l'administration, voire comme langue véhiculaire, dans l'affichage, dans l'enseignement, dans l'économie.

Le tagalog est une langue malayo-polynésienne, famille qui couvre une aire géographique très vaste, de Madagascar à Formose en passant par l'Indonésie et les Philippines, et rassemble plus de 300 millions de locuteurs. Comme les autres langues de cette famille, le tagalog se caractérise par un grand nombre (plusieurs dizaines) de préfixes, infixes (qui s'intercalent au milieu du mot, comme dans le « javanais », langage secret et humoristique des étudiants français) et suffixes qui viennent compléter une base nominale et permettent de composer de nouvelles unités et d'exprimer de nombreuses nuances.

L'espagnol a presque complètement disparu des usages linguistiques des Philippines, mais le vocabulaire du tagalog reste fortement imprégné de mots, adaptés, qu'il en a conservés : *kumesta?* (comment ça va?) < *como esta?*...

## \* Tamoul

தமிழ்

### Famille

langue dravidienne

### Localisation

Inde (Etat du Tamil Nadu), Sri Lanka, Malaisie

### Nombre de locuteurs

environ 66 millions

### Système d'écriture

alphabet tamoul

Cette langue dravidienne (*cf.* malayalam) est une des langues constitutionnelles de l'Inde et la langue nationale de l'Etat du Tamil Nadu (sud). Elle est également fortement présente à Sri Lanka (environ 3 millions de locuteurs) où elle est parlée par les populations tamoules en conflit depuis plusieurs années avec la majorité singhalaise (langue indo-européenne). Entre autre en raison de ce conflit, le tamoul est aussi parlé par une très importante diaspora : à Singapour (où il est langue officielle), en Malaisie (plus d'un million de locuteurs), en Afrique, en Europe...

Le système d'écriture tamoul semble moins directement lié à l'alphabet pali (utilisé pour l'écriture des textes bouddhiques sacrés) que les autres langues du sud de l'Inde. Phonétiquement, la langue tamoule comporte quelques sons qui nous sont inconnus (dont une sorte de l prononcé en repliant la langue mais sans qu'elle ne touche le palais) et qui s'avèrent par conséquent difficiles à percevoir.

Le tamoul est une langue à huit cas, utilisant de nombreux suffixes ; le verbe est placé à la fin de la phrase. A noter, comme en français, que le pluriel est utilisé pour les formes de politesse. Et, comme dans les autres langues dravidiennes, le tamoul possède deux formes pour le pronom de 1<sup>re</sup> personne du pluriel : un nous *inclusif* et un nous *exclusif*.

Le vocabulaire contient de nombreux emprunts au sanskrit, qui viennent souvent doubler un mot d'origine dravidienne et sont surtout utilisés par la caste la plus élevée, celle des brahmanes.

## Tchèque – *čeština*

### Famille

langue slave (famille indo-européenne)

### Localisation

Tchéquie (minorités dans les pays voisins : Pologne, Autriche, ...)

### Nombre de locuteurs

environ 11 millions

### Système d'écriture

alphabet latin

Langue nationale de la Tchéquie, le tchèque partage les principales caractéristiques des langues slaves (système de déclinaison à sept cas... *cf.* russe). Elle est proche en particulier du slovaque, autre langue de l'ancienne Tchécoslovaquie, avec laquelle la compréhension ne pose pour ainsi dire pas problème.

Comme les autres langues slaves des pays qui ne sont pas de religion orthodoxe, le tchèque utilise un alphabet latin aménagé et augmenté de quelques lettres (č, š, ž; *cf.* slovaque). Phonétiquement, il possède entre autres une sorte de [k] grasseyé et prononcé très en arrière de la gorge ainsi qu'une lettre ř, prononcée approximativement entre r «roulé» et [ʒ] (*cf.* la prononciation fréquente «Dvorjak» du nom du musicien célèbre qui s'écrit Dvořák).

## Tibétain

བོད་སྐད་

### Famille

langue sino-tibétaine (groupe tibéto-birman)

### Localisation

Tibet et deux autres provinces de Chine, Bhoutan, Népal, Inde

### Nombre de locuteurs

environ 1 million et demi

### Système d'écriture

alphabet tibétain (dérivé du devanagari)

Le tibétain, sous des formes diverses (variétés dialectales ou langues différentes appartenant au même groupe), est parlé dans plusieurs pays : au Tibet, rattaché à la Chine, dans d'autres provinces de Chine, en Inde (Ladakh, Sikkim) où l'on trouve de nombreux réfugiés provenant du Tibet, au Bhoutan, qui est le seul Etat indépendant de langue majoritairement tibétaine. Enfin, des groupes importants de Tibétains vivent au Népal (les *Newar* dans la vallée de Katmandou, les *Sher-pa*, ce qui signifie littéralement les «gens de l'est»). Toutes ces populations parlent donc des langues tibétaines, mais pas la *même* langue : il y a plus de 50 langues tibétaines entre lesquelles la communication n'est pas impossible, mais pas aisée non plus.

L'écriture tibétaine est dérivée de l'alphabet devanagari (sanskrit), quoique certaines lettres se prononcent différemment. Elle a été formée au VII<sup>e</sup> siècle, lorsque le bouddhisme s'est développé dans le pays. L'écriture tibétaine présente en outre la particularité de marquer par un point la fin de chaque syllabe. Le tibétain est une langue monosyllabique à tons, comme le chinois ; mais il ne possède que trois tons, distingués par leur hauteur (haut, moyen, bas), qui ont une importance bien moindre qu'en chinois. Comme le chinois, le tibétain différencie peu les mots en catégories telles que nom, adjectif ou verbe. Les verbes ne sont pas conjugués mais utilisent des auxiliaires pour marquer les temps : le mot *finir* marque le passé, *devant* marque le futur. Le genre des noms n'est pas marqué, sauf si nécessaire. Le tibétain contient en revanche un système important de «mots honorifiques» qui varient selon le statut de la personne à laquelle on s'adresse.

## \* Tigrinia

ትግርኛ

### Famille

langue sémitique (même famille que l'arabe, l'hébreu, l'amharique, etc.)

### Localisation

Erythrée, Ethiopie

### Nombre de locuteurs

environ 5 millions

### Système d'écriture

alphabet éthiopien (s'écrit de gauche à droite)



Le tigrinia (ou tigrigna) fait partie de la branche éthiopienne de la famille sémitique ; l'amharique et le tigréen en sont de proches parents. Leur ancêtre commun est le *guèze*, langue littéraire classique d'Éthiopie, encore utilisé comme langue religieuse. Ces langues s'écrivent avec un alphabet propre, parent des anciennes écritures sud-arabiques (aujourd'hui disparues) et fondé sur le même principe qu'utilisent les alphabets indiens (signes complémentaires attachés aux consonnes pour marquer les voyelles). A l'origine, l'alphabet éthiopien s'écrivait de droite à gauche, puis il changea de sens, probablement sous l'influence des Grecs. Deux points sont placés après chaque mot afin de marquer la séparation d'avec le mot suivant.

### \* Turc – türkçe

#### Famille

langue du groupe turco-mongol ou altaïque (famille ouralo-altaïque)

#### Localisation

Turquie, Chypre, mais aussi minorités en Bulgarie, Macédoine, Ouzbékistan, ...

#### Nombre de locuteurs

environ 60 millions

#### Système d'écriture

alphabet latin

Le turc appartient à la branche altaïque des langues du groupe ouralo-altaïque, qui doit son nom aux monts Altaï, un massif d'Asie centrale s'étendant des frontières du Sud de la Russie à la Mongolie et à la Chine. La famille des langues altaïques regroupe une cinquantaine de langues, dont les langues de la sous-famille turque (*turc*, *turkmène*, *ouzbek*, *azéri*, *kirghiz*, ...), celles du groupe mongol et celles du groupe toungouze qui sont parlées jusqu'aux confins de la Sibérie et de la Chine. Parmi toutes les langues altaïques, c'est le turc qui compte le plus grand nombre de locuteurs (environ 60 millions).

A l'origine, le turc s'écrivait avec l'alphabet arabe. En 1928, le président Mustafa Kemal Atatürk introduisit un programme de laïcisation et d'occidentalisation du pays ; depuis lors, le turc s'écrit avec l'alphabet latin (auquel est ajouté une lettre *ı*, sans point, prononcé entre [i] et [e]).

Le turc est caractérisé par une importante suffixation : les mots sont composés de longues séquences d'unités, chacune remplissant une fonction grammaticale précise. Une autre particularité est le phénomène de l'harmonie vocalique : les voyelles des différentes unités d'un mot sont changées en fonction des voyelles de la racine du mot ; par exemple le pluriel de *ev* (maison) est *evler*, alors que le pluriel de *at* (cheval) devient *atlar*. Les noms n'ont pas de genre (masculin ou féminin) et il n'y a pas d'article défini. De façon générale, la grammaire est très régulière et il n'y a pratiquement pas d'exceptions.

### \* Vietnamien

### Tiếng Việt

#### Famille

langue môn-khmer de la famille austro-asiatique (mais le vietnamien a intégré de nombreux traits et mots du chinois)

#### Localisation

Vietnam (importantes communautés migrantes dans de nombreux pays)

#### Nombre de locuteurs

environ 68 millions de locuteurs

#### Système d'écriture

alphabet latin complété par quelques signes, entre autres pour indiquer les tons

Le vietnamien, parlé au Vietnam par environ 85% de la population, est la deuxième langue en importance des langues tonales d'Asie après le chinois. L'origine du vietnamien n'est pas clairement définie. C'est une langue très composite, dans laquelle on trouve une influence antérieure très profonde des langues môn-khmer, qui ne sont pas des langues tonales, ainsi que des langues thaï ; mais l'influence culturelle du chinois y a laissé une très forte empreinte dont près de la moitié du vocabulaire ; jadis d'ailleurs, jusqu'au XIII<sup>e</sup> siècle, le vietnamien s'écrivait au moyen des idéogrammes chinois.

Aujourd'hui, les caractères latins ont été adoptés. Mais le système de transcription du vietnamien est considérablement différent du nôtre, d'une part en raison des signes diacritiques qui permettent de noter les tons et certaines voyelles particulières, d'autre part du fait que certaines lettres n'y ont pas la même valeur qu'en français (*tr*, par exemple, se lit [tʃ], *ph* se lit [f], *x* se lit [s]...).

Le vietnamien est une langue monosyllabique, assez homogène : il existe principalement deux dialectes, celui du Nord qui a six tons et celui du Sud qui n'en a que cinq.

### Wolof – wolof

#### Famille

langue nigéro-congolaise

#### Localisation

Sénégal, Gambie

#### Nombre de locuteurs

environ 3 millions (environ 7 millions en tant que langue première ou seconde)

#### Système d'écriture

alphabet latin et, parfois, alphabet arabe (car la population est en majeure partie musulmane)

Le wolof est la principale langue véhiculaire du Sénégal. Il est la langue maternelle de 36% des Sénégalais, mais près de 80% de la population le parle ou le comprend. A côté du français, qui conserve un statut privilégié dans le système éducatif et dans l'administration, le wolof fait partie des six langues reconnues comme langues nationales : le *wolof*, le *peul*, le *sérère*, le *diola* (qui appartiennent toutes à la famille nigéro-congolaise), le *malinké* et le *soninké* qui sont des langues du groupe mandé.

Dès 1960, le chef d'Etat Léopold Sédar Senghor (qui était linguiste et poète) a lancé d'importants travaux d'élaboration d'alphabets officiels, basés sur l'alphabet latin, et de terminologies pour ces six langues. Puis (dès le début des années 70) il instaura une politique d'éducation bilingue comprenant le français d'une part et l'une ou l'autre des six langues nationales d'autre part. Toutefois, dans la pratique, cette éducation bilingue peine à se mettre en place et le français reste encore largement prédominant.

Le wolof est une langue à classes nominales (*wo-* est par exemple un préfixe de classe marquant le pluriel des noms de personnes ; cf. langues bantoues). Son système de numération est de type quinaire (base 5), comme celui du nahuatl par exemple.