

OUVRAGE COORDONNÉ
PAR **LE RÉSEAU MAAYA**

NET. LANG

RÉUSSIR LE CYBERESPACE MULTILINGUE





NET.LANG EST DISPONIBLE EN PLUSIEURS LANGUES ET SUR PLUSIEURS SUPPORTS :

En version imprimée (français ou anglais)

480 pages, format 17 × 22,5 cm

prix : 34 euros

en français : ISBN 978-2-915825-08-4

en anglais : ISBN 978-2-915825-09-1

En librairie ou sur <http://cfeditions.com>

Au format ePub (français ou anglais)

sans DRM

en français : ISBN 978-2-915825-25-1

en anglais : ISBN 978-2-915825-26-8

<http://net-lang.net>

En pdf (français ou anglais)

sans DRM

en français : ISBN 978-2-915825-23-7

en anglais : ISBN 978-2-915825-24-4

<http://net-lang.net>

D'AUTRES VERSIONS ET TRADUCTIONS
[HTTP://NET-LANG.NET](http://NET-LANG.NET)

NET.LANG

RÉUSSIR LE CYBERESPACE MULTILINGUE

Site compagnon multilingue :

<http://net-lang.net>

Chez le même éditeur :

Libres savoirs,

les biens communs de la connaissance

Ouvrage collectif coordonné par l'association VECAM

Mai 2011, ISBN 978-2-915825-06-0

Aux sources de l'utopie numérique

De la contre-culture à la cyberculture : Steward Brand, un homme d'influence

Par Fred Turner

Avril 2012, ISBN 2-915825-10-6

Dans le labyrinthe

Évaluer l'information sur internet

Par Alexandre Serres

Avril 2012, ISBN 978-2-915825-22-0

Catalogue complet et vente en ligne :

<http://cfeditions.com>

ISBN édition PDF 978-2-915825-23-7

C&F éditions, mars 2012

35C rue des Rosiers – 14000 Caen

<http://cfeditions.com>

L'ouvrage est publié sous licence Creative Commons : attribution, partage à l'identique (<http://creativecommons.org/licenses/by-sa/3.0/fr/>).

RÉSEAU MAAYA

NET.LANG

RÉUSSIR LE CYBERESPACE MULTILINGUE

ÉDITION ET COORDINATION :
LAURENT VANNINI
HERVÉ LE CROSNIER

C&F ÉDITIONS
2012

RÉSEAU MAAVA

NET.LANG

RÉUSSIR LE CYBERESPACE MULTILINGUE

ÉDITION ET COORDINATION :
LAURENT VANNINI
HERVÉ LE GROSNIER

2013
G&F ÉDITIONS

TABLE DES MATIÈRES

Générique 9

AVANT-PROPOS 10

Irina Bokova *Directrice générale de l'UNESCO* 13

Abdou Diouf *Secrétaire général de la Francophonie* 17

José Luis Dicenta *Secrétaire général de l'Union Latine* 21

Dwayne Bailey *Directeur de recherche de ANLoc* 23

Daniel Prado *Secrétaire exécutif du Réseau Maaya* 27

PARTIE 1 32

QUAND LES TECHNOLOGIES RENCONTRENT LE MULTILINGUISME

Daniel Prado
Présence des langues dans le monde réel et le cyberspace 35

Michaël Oustinoff
L'anglais ne sera pas la *lingua franca* de l'Internet 55

Éric Poncet
Innovation technologique et maintien des langues 73

Maik Gibson
Préserver l'héritage des langues éteintes ou en danger 79

Marcel Diki-Kidiri
Le cyberspace dans l'éducation en langue maternelle 95

PARTIE 2 **108** **ESPACES NUMÉRIQUES**

Stéphane Bortzmeyer	
Le multilinguisme dans la normalisation de l'internet	111
Mikami Yoshiki & Shigeaki Kodama	
Mesurer la diversité linguistique sur le Web	125
Joseph Mariani	
Les technologies de la langue en soutien au multilinguisme	149
Vassili Rivron	
L'usage de Facebook chez les Éton du Cameroun	171
Pann Yu Mon & Madhukara Phatak	
Moteurs de recherche et langues asiatiques	179
Hervé Le Crosnier	
Bibliothèques numériques	197
Dwayne Bailey	
Localisation des logiciels : open source et multilinguisme numérique	219
Mélanie Dulong De Rosnay	
Traduction et localisation des licences Creative Commons	239

PARTIE 3 **247** **LE MULTILINGUISME NUMÉRIQUE** **POUR CONSTRUIRE DES** **SOCIÉTÉS INCLUSIVES**

Viola Krebs & Vicent Climent-Ferrando	
Langues, cyberspace et migrations	249
Annelies Braffort & Patrice Dalle	
Accessibilité du cyberspace : langues des signes	269
Tjeerd de Graaf	
Intérêt des archives orales pour les langues menacées	291
Evgeny Kuzmin	
Politiques linguistiques pour contrer la marginalisation des langues	309

Tunde Adegbola
Multimédia, langues des signes,
langues écrites et orales 337

Adel El Zaim
Cyberactivisme et langues régionales
dans les printemps arabes de 2011 351

Adama Samassékou
Multilinguisme, Objectifs du Millénaire
pour le développement et le cyberspace 363

PARTIE 4 376 **LE MULTILINGUISME SUR INTERNET, UNE QUESTION MULTILATÉRALE**

Isabella Pierangeli Borletti
Décrire le monde : Multilinguisme,
Internet et Droits de l'Homme 379

Stéphane Bortzmeyer
La gouvernance de l'internet
à l'heure du multilinguisme 405

Marcel Diki-Kidiri
Principes éthiques requis pour une présence équitable
des langues dans la société de l'information 419

Stéphane Grumbach
L'Internet en Chine 435

Michaël Oustinoff
L'économie des langues 441

Daniel Prado & Daniel Pimienta
Quelles politiques publiques
pour les langues dans le cyberspace ? 457

CONCLUSION 471 **L'AVENIR SE PARLE, S'ÉCRIT ET SE LIT DANS TOUTES LES LANGUES**

Adama Samassékou *président de Maaya*



LAURA KRAFTOWITZ née à Pittsburgh en 1982, partage dorénavant son temps entre Paris et le Moyen-Orient. Elle a écrit de nombreux articles sur le conflit israélo-palestinien, et un livre à paraître *The End of Abu Jameel Street*, appuyé sur son expérience dans la bande de Gaza. Elle termine actuellement son Master 2 en Études politiques et Histoire à l'École des Hautes Études en Sciences Sociales à Paris, où elle mène des recherches sur l'idéation binationale israélienne et palestinienne au cours du dernier siècle.



HERVÉ LE CROSNIER est maître de conférences à l'Université de Caen Basse-Normandie, où il enseigne les technologies de l'internet et la culture numérique. Il est actuellement en délégation auprès de l'ISCC (*Institut des Sciences de la Communication du CNRS*). Sa recherche porte sur l'impact de l'internet sur l'organisation sociale et culturelle, et l'extension du domaine des biens communs de la connaissance. Hervé Le Crosnier est un des fondateurs de C&F éditions.



JOHN ROSBOTTOM John Rosbottom était maître de conférence en Informatique à l'Université de Portsmouth, Royaume-Uni, jusqu'à sa retraite en 2010. Il réalise maintenant les projets ajournés au fil d'une vie de travail universitaire bien remplie. Notamment des projets informatiques et des traductions de livres et de documents techniques.



LAURENT VANNINI après avoir été activiste de la diversité linguistique dans le réseau Babels, journaliste, puis consultant pour le secteur des télécommunications, a décidé de parcourir le monde et de reprendre des études en « Arts et Langages » à l'EHESS. Ses travaux portent actuellement sur « la disparition de l'animal et de l'oubli ». En marge de son activité éditoriale liée aux TIC, il a traduit le livre *Aux sources de l'utopie numérique: de la contre-culture à la cyberculture*, Steward Brand, un homme d'influence de Fred Turner (C&F éditions, 2012).

GÉNÉRIQUE

Conseil scientifique Maaya

Adama Samassékou
Daniel Prado
Daniel Pimienta
Marcel Diki-Kidiri
Louis Pouzin
Yoshiki Mikami
Evgeny Kuzmin

Édition et coordination

Laurent Vannini
Hervé Le Crosnier

Traduction

John Rosbottom
Laura Kraftowitz
Laurent Vannini
Francisca Cabrera
Alexandre Mussely

Conception graphique et mise en page

Nicolas Taffin
Kathleen Ponsard

Partenaires

Unesco, secteur de la communication et de l'information
La Francophonie
Union Latine
ANLoc, Afrique du Sud
IDRC / CRDI, Canada

AVANT- PROPOS



Organisation
des Nations Unies
pour l'éducation,
la science et la culture

En coopération avec le
**Secteur de
la communication
et de l'information**

L'UNESCO s'emploie à créer les conditions d'un dialogue entre les civilisations, les cultures et les peuples, fondé sur le respect de valeurs partagées par tous. C'est par ce dialogue que le monde peut parvenir à des conceptions globales du développement durable intégrant le respect des droits de l'homme, le respect mutuel et la réduction de la pauvreté, tous ces points étant au coeur de l'action de l'UNESCO.

La mission de l'UNESCO est de contribuer à l'édification de la paix, à l'élimination de la pauvreté, au développement durable et au dialogue interculturel par l'éducation, les sciences, la culture, la communication et l'information.

<http://unesco.org>

PRÉSERVER ET PROMOUVOIR LA DIVERSITÉ LINGUISTIQUE

**PAR IRINA BOKOVA,
Directrice générale de l'UNESCO**

Les langues sont des composantes essentielles du patrimoine commun de l'humanité tout en appartenant à chaque être humain. Elles sont le véhicule premier et principal de l'expression des identités, de la communication des idées, de l'apprentissage d'une autonomie pédagogique, économique et politique, ainsi que de la promotion d'un développement humain durable et de la paix dans le monde.

Les langues sont fondamentales pour échanger information et connaissance, et pour transmettre des sagesses culturelles uniques, tant entre générations qu'entre nations. Elles constituent une partie essentielle de l'identité des individus et des peuples, et elles sont d'une importance vitale dans la préservation de la diversité culturelle à l'échelle mondiale. Elles ouvrent la porte au dialogue, à la coopération et à la compréhension mutuelle. Dans une telle perspective, un espace linguistique de diversité et de pluralité contribue à l'expansion des conditions d'existence d'un tel dialogue, en permettant à tout individu sans exception d'y participer librement dans les langues de son choix.

Dans le même temps, les langues sont des ressources fragiles qui exigent qu'on les protège, qu'on les revitalise et qu'on en fasse la promotion. Aujourd'hui, un nombre considérable de langues risquent de s'éteindre. Quelque 97 % de la population mondiale parle environ 4 % des langues de la planète. Inversement, approximativement 96 % des langues existantes ne sont parlées que par 3 % des êtres humains à travers le monde. Cela signifie qu'environ la moitié des 6000 langues composant la diversité

linguistique mondiale voient leur nombre de locuteurs diminuer. On estime qu'environ 90 % de la totalité des langues actuelles seront remplacés par l'une ou l'autre des langues dominantes avant la fin du XXI^e siècle.

Alors que les Technologies de l'Information et de la Communication (TIC) sont désormais au cœur de la vie sociale, culturelle, politique et économique, il est essentiel de s'assurer que chaque être humain ait accès et puisse contribuer avec des contenus qui lui soient propres à un internet multilingue. Les TIC sont des outils puissants pour la préservation et la promotion de la diversité linguistique. En principe, l'internet est ouvert à toutes les langues du monde, mais cela ne devient une réalité qu'à certaines conditions, et notamment lorsque les ressources humaines et financières nécessaires sont en place. Un internet multilingue est essentiel pour que les nations, les communautés et les individus accèdent, partagent et utilisent l'information et les ressources qui sont nécessaires au développement durable et à la gestion de l'innovation et du changement.

L'UNESCO s'est fermement engagé dans la promotion du multilinguisme sur l'internet. C'est au travers d'un cyberspace permettant l'expression plurielle des langues que nous pourrions mettre en commun toute la richesse contenue dans la diversité. Ces objectifs guident l'Organisation dans son travail avec l'Internet Corporation of Assigned Names and Numbers (ICANN). Ils font également l'objet de nombreuses discussions au sein du Forum sur la Gouvernance de l'Internet (FGI), ainsi que dans les forums du Sommet Mondial sur la Société de l'Information (SMSI). Le partenariat entre l'UNESCO et l'Union Internationale des Télécommunications (UIT) à l'intérieur de la Commission sur le Haut-débit pour le Développement numérique a également mis l'accent sur la nécessité d'une grande diversité linguistique et culturelle des contenus et logiciels locaux; c'est un objectif clé pour tous les décideurs du monde qui sont engagés dans l'établissement du haut-débit pour tous.

De même, l'importance de la diversité culturelle et linguistique trouve un écho dans la *Recommandation sur la promotion et l'usage du multilinguisme et l'accès universel au cyberspace*, adoptée par la Conférence générale de l'UNESCO en 2003. Cette Recommandation appelle à l'adoption de mesures urgentes et concrètes pour promouvoir la diversité linguistique, en développant un internet multilingue qui assure la préservation des langues, notamment celles en voie d'extinction. Elle encourage le développement de projets pilotes, de ressources et d'outils de gestion des

contenus multilingues. Elle propose également l'établissement d'un accès plus large et plus équitable aux services et réseaux d'information, tout en réaffirmant la nécessité d'un équilibre entre les intérêts des détenteurs de droits et ceux du public.

Depuis dix ans l'UNESCO promeut le concept de sociétés de la connaissance ouvertes, pluralistes, équitables et participatives. Internet et les réseaux sociaux numériques ont un rôle fondamental à jouer dans l'éclosion de telles sociétés inclusives. L'internet deviendra une telle plateforme ouverte et équitable, s'il est guidé par les principes d'ouverture, de liberté d'expression, de diversité culturelle et de multilinguisme.



ORGANISATION INTERNATIONALE DE LA FRANCOPHONIE

Forte d'une population de plus de 890 millions d'habitants et de 220 millions de locuteurs de français de par le monde, l'Organisation internationale de la Francophonie (OIF) a pour mission de donner corps à une solidarité active entre les 75 États et gouvernements qui la composent (56 membres et 19 observateurs) –soit plus du tiers des États membres des Nations Unies.

La Francophonie agit dans le respect de la diversité culturelle et linguistique et au service de la promotion de la langue française, de la paix et du développement durable.

<http://www.francophonie.org/>

PROMOUVOIR LA DIVERSITÉ CULTURELLE DANS L'UNIVERS NUMÉRIQUE

PAR ABDOU DIOUF,
Secrétaire général de la Francophonie

Une grande partie de notre destin commun se joue au sein de la société de l'information qui, du Sud au Nord, de l'Est à l'Ouest, est en train de bouleverser et de façonner, non seulement nos modes d'information, de communication, de production et de consommation, mais aussi, nos modes de réflexion, de création et d'accès aux savoirs, nous ouvrant à tous, particulièrement à notre jeunesse, des perspectives inédites.

Face aux grands enjeux planétaires, la Francophonie, une organisation internationale regroupant 75 États et gouvernements membres et observateurs, soit plus du tiers des membres des Nations Unies, a très tôt saisi la nécessité de se mobiliser, de se coordonner et de relever les défis de la société de l'information. Sa présence active et concertée au sein des instances internationales de décision lui permet de donner résonance à ses positions, de se constituer en force de propositions, de peser de tout son poids pour infléchir les décisions internationales dans le sens des aspirations de ses États et gouvernements membres.

Aujourd'hui, l'appropriation de la culture numérique favorise un intense bouillonnement créatif dans tous les champs de l'activité humaine. Le passage au numérique est un risque autant qu'une chance pour le français comme pour les autres expressions linguistiques au sein de la Toile. La question du poids respectif des langues sous tous les formats en circulation dans les réseaux est d'ailleurs très mal documentée. Depuis 2002, à travers notamment sa Direction de la Francophonie numérique et son Observatoire de la Langue française, notre organisation a soutenu les études menées régulièrement par Funredes sur la place du français sur internet. Prenant en compte les évolutions constantes du cyberspace, l'Organisation Internationale de la Francophonie apporte également son appui à toutes les initiatives en cours qui permettront d'établir de nouveaux indicateurs de mesure de la diversité linguistique dans l'univers numérique.

L'engagement francophone vise à garantir les conditions préalables à l'existence de la diversité culturelle et de la pluralité linguistique dans la

société de l'information. L'objectif est que la communauté francophone puisse exprimer ses singularités et ainsi s'approprier la culture numérique dans toute sa diversité. Dans ce contexte, le libre accès aux contenus numériques et aux technologies innovantes est une priorité.

Dans le domaine de la culture comme dans celui du numérique, la Francophonie entend lutter contre l'uniformisation et préserver les valeurs dans lesquelles une société trouve, en même temps que son identité, les ferments de son développement ou de son renouveau.

Le soutien apporté par l'OIF à la présente publication s'inscrit dans cette dynamique et témoigne également des synergies en matière de diversité culturelle et linguistique que la Francophonie cultive depuis de nombreuses années avec plusieurs organismes intergouvernementaux et acteurs internationaux de la société civile, notamment l'Unesco, le réseau Maaya et l'Union Latine.



UNION LATINE

Pour une latinité plurielle

Les peuples de langues romanes sont d'origines très diverses ; ils ont cependant en partage le même patrimoine linguistique et le même système de références historiques, juridiques et culturelles. Il est donc naturel que cette famille, même dispersée et très élargie, se dote d'une institution vouée à la promotion et à la diffusion de l'héritage commun et des identités du monde latin. Cette action est d'autant plus nécessaire à l'heure où la préservation de la diversité culturelle constitue une des préoccupations majeures du monde contemporain.

<http://www.unilat.org/>

FAVORISER LA PRÉSENCE DES LANGUES DANS LE CYBERESPACE

PAR L'AMBASSADEUR JOSÉ LUIS DICENTA,
Secrétaire général de l'Union Latine

L'Union Latine est une Organisation intergouvernementale, créée il y a cinquante-sept ans, qui a mené une action significative tout au long de ces années, aussi bien dans le domaine culturel (rencontres, protection du patrimoine, festivals de cinéma, expositions, etc.) que linguistique.

L'Union Latine a su devenir dans ce dernier secteur un organisme pionnier dans la mesure des langues dans le cyberspace et demeure aujourd'hui l'une des rares entités au monde qui se consacre à cette tâche. Une tâche dont on ne peut nier l'importance étant donnés les paramètres de notre monde contemporain, où le cyberspace constitue un véritable territoire qu'il convient de connaître et comprendre.

Par ailleurs, l'Union Latine, soucieuse d'enrichir les terminologies aussi bien romanes que d'autres langues non-latines mais néanmoins parlées dans ses pays membres, contribue actuellement à la création de réseaux, d'associations et d'entités ayant pour objectifs la revitalisation et la modernisation de ces langues.

L'intercompréhension entre les langues romanes constitue l'une des autres avancées de l'Union Latine qui a notamment organisé des rencontres et des séminaires de formation de formateurs susceptibles de promouvoir, à l'avenir, cette discipline indispensable pour faciliter en particulier la compréhension entre les populations de pays limitrophes.

La promotion de politiques de soutien au développement des langues est également un objectif de travail de l'Union Latine qui a contribué à créer des indicateurs permettant l'orientation de politiques spécifiques au secteur linguistique.

Par le soutien qu'elle apporte à la réalisation de *Net.lang*, l'Union Latine entend ainsi favoriser la modernisation de toute langue dont la présence dans le cyberspace s'avère insuffisante, en offrant aux décideurs un guide pratique de décisions efficaces. C'est pourquoi nous nous sommes engagés avec conviction dans ce projet et que nous lui souhaitons un grand succès.



ANLOC
The African Network for Localisation
Réseau africain pour la Localisation

ANLoc est un réseau pan-africain de chercheurs des universités, d'acteurs de la société civile et du secteur privé. Ces partenaires provenant de plus de neuf pays s'assistent mutuellement afin de couvrir les besoins linguistiques des langues locales. Pour l'ANLoc, chaque africain doit pouvoir participer à l'univers numérique. Il y a plus d'un milliard d'africains, et nous souhaitons que leurs voix soient entendues de par le monde.

<http://www.africanlocalisation.net/>



IDRC - CRDI
International development
research centre, Canada
Centre de recherches pour
le développement international, Canada

Le CRDI, qui est un élément clé de l'aide internationale du Canada depuis 1970, appuie des travaux de recherche dans les pays en développement afin d'y favoriser la croissance et le développement. Il veille aussi à promouvoir la diffusion des connaissances ainsi acquises auprès de responsables des politiques, de chercheurs et de collectivités de par le monde. Il en résulte des solutions locales, novatrices et durables, qui offrent des choix aux personnes qui en ont le plus besoin et font changer les choses.

<http://www.idrc.ca/>

LES LANGUES LIBÈRENT LE POTENTIEL HUMAIN

PAR DWAYNE BAILEY,

Directeur de recherche, African Network for Localisation

Le Réseau africain pour la localisation – ANLoc, que l'on prononce *Unlock* [déverrouiller – Ndt] – a choisi ce nom pour souligner combien le vaste potentiel de l'humanité restait enfermé sous l'emprise du monolinguisme. Rien d'étonnant donc à ce que nous ayons participé avec enthousiasme à ce livre. Comme tous les internautes, nous nous réjouissons des innovations mises en œuvre sur l'internet et dans le monde numérique, mais nous sommes plus encore mobilisés quand de telles innovations sont disponibles pour toutes les langues.

Les actions menées par l'ANLoc, ainsi que notre soutien à ce livre, sont rendus possibles par le Centre de Recherche pour le Développement International (IDRC/CRDI) du Canada, qui épaulé financièrement une grande partie de notre travail pour ouvrir l'informatique à toutes les langues d'Afrique.

Nous nous sommes principalement concentrés sur les aspects technologiques du multilinguisme. Notre travail nous a permis de prendre conscience que la majorité des acteurs ne se préoccupaient pas du multilinguisme, effrayés qu'ils étaient par sa complexité technique. Les contenus monolingues ainsi produits ne se résument pas aux traditionnels sites web entièrement en anglais. En Afrique, le monolinguisme s'exprime aussi en français, en portugais ou en arabe. Le monolinguisme, c'est le choix exclusif des langues officielles au détriment des langues nationales ou minoritaires.

Faire évoluer une technologie bloquée dans le monolinguisme demande des efforts. Mais nous avons découvert qu'au regard de l'impact généré, un grand nombre de préalables sont finalement très simples, et s'acquièrent sans peine. Des choses aussi élémentaires que l'ajout de caractères à une police typographique, la fabrication d'un clavier ou la traduction d'un logiciel open source, rendent possible d'écrire et de lire du contenu numérique dans toutes les langues. Certaines de ces interventions nécessitent un solide niveau d'expertise informatique. L'ANLoc s'efforce de les prendre à son compte, pour rendre les usagers autonomes, afin de laisser aux spécialistes des langues le soin de se concentrer sur leur art.

Ces interventions de l'ANLoc n'ont pas modifié la face du monde, mais nous avons levé un nombre important de barrières techniques qui empêchaient toute participation à l'univers numérique des locuteurs ne connaissant que leur langue maternelle. Quand leur communauté dispose des moyens de leur autonomie, ces personnes ont les moyens de construire un internet qui leur soit adapté. Nous avons simplement levé les obstacles techniques inhérents aux insuffisances de la technologie en regard du traitement des langues.

Le futur d'une langue est entre les mains de celles et ceux qui la parlent. Cela reste vrai à l'ère numérique comme à toute autre époque. Une fois que ces barrières techniques ont été levées, un linguiste ou un programmeur informatique ne peuvent en aucun cas forcer qui que ce soit à participer au monde numérique. Heureusement, il ne semble pas que quiconque ait besoin d'être forcé pour utiliser Google, Facebook ou Twitter. L'ANLoc s'est penché sur cette profusion de potentiel humain, et nous avons travaillé d'arrache-pied pour comprendre comment nous pouvions valoriser l'énergie de tous ces volontaires. Une partie de notre travail consiste à créer les technologies web permettant aux communautés de mettre en place les traductions communautaires coopératives. Ceci a permis à ces communautés de bâtir une terminologie adaptée aux Technologies de l'Information et de la Communication dans leur langue, ainsi que de traduire Firefox dans dix langues d'Afrique. Cette tendance nouvelle à une information traduite « par les foules » dans leur langue maternelle aboutit à une véritable dynamique de participation. Des bénévoles traduisent des logiciels, des sites de réseau social, des moteurs de recherche, des articles de Wikipédia ou rédigent des correcteurs orthographiques. L'ère numérique pourrait bien se révéler celle du triomphe pour toutes les langues.

L'ANLoc a contribué à relever les nombreux défis posés par le multilinguisme. Le livre *Net.lang* nous enthousiasme parce qu'il porte auprès des décideurs politiques et économiques, le message qui est le nôtre, et que nous partageons avec tous les contributeurs, à savoir que le multilinguisme est non seulement possible, mais qu'il est extrêmement puissant. Appuyé sur le travail d'experts internationaux et fort du soutien de l'Unesco, nous pensons que cet ouvrage agira comme un catalyseur pour inciter au développement d'un cyberspace multilingue, pour le bénéfice de tous. Pour libérer le potentiel humain enchaîné par un monde monolingue.



Maaya / Réseau mondial pour la diversité linguistique est un réseau multi-acteurs créé dans le but de contribuer à la valorisation et la promotion de la diversité linguistique dans le monde. Terme de la langue bambara, Maaya pourrait se traduire par le néologisme « Humanitude ». Le Réseau Maaya a été créé dans le cadre du Sommet Mondial sur la Société de l'Information (SMSI), durant lequel la diversité culturelle et linguistique dans le cyberspace a été identifiée comme l'une des priorités. Maaya a été initiée par l'Académie Africaine des Langues (ACALAN), sous les auspices de l'Union Africaine.

<http://www.maayajo.org/>

PETIT RETOUR SUR LA GENÈSE DE NET.LANG

PAR DANIEL PRADO,
Secrétaire exécutif du Réseau Maaya

Le livre *Net.lang* que vous avez actuellement entre les mains, en version imprimée ou bien dans une des versions numériques (en pdf, en livre numérique au format ePub, ou par l'accès direct au site web) a été imaginé et coordonné par le Réseau Maaya.

Réseau mondial pour la diversité linguistique, Maaya est né dans le cadre du Sommet mondial sur la Société de l'Information (SMSI). Ses objectifs portent sur la réalisation d'un monde réellement multilingue, où chaque langue, et les richesses de culture et de connaissance qu'elle porte, ait droit de cité, recueille le respect et participe à l'expansion de la société des savoirs partagés. Un tel objectif passe par la mobilisation des individus et des associations, mais aussi par la prise en compte du multilinguisme par les gouvernements et les institutions. C'est dans ce but que Maaya a proposé l'écriture collective d'un ouvrage permettant à toutes les personnes et structures concernées d'agir pour développer un cyberspace multilingue.

Nous sommes conscients que le cyberspace présente à la fois une menace et une opportunité pour le multilinguisme. Menace, car les langues les plus outillées et celles parlées dans des États dominants s'imposent devant les autres et trouvent appui dans la technicité même de ce réseau ; opportunité, car, par sa facilité d'accès et son universalité, le cyberspace permet de donner la parole à ceux qui ne l'avaient pas au travers des autres outils d'enregistrement et de diffusion du savoir. Nous croyons que cette facilité d'accès, la capacité du réseau internet a mobiliser et coordonner de nombreux individus et sa puissance multimédia va permettre d'assurer ainsi le sauvetage et la revitalisation des langues minoritaires.

Net.lang a donc la prétention de donner à un grand public concerné et éclairé des pistes et des analyses pour que soit tiré le meilleur bénéfice de l'internet afin de permettre une représentation équitable des langues.

Ouvrage de vulgarisation, prétendant présenter de façon exhaustive, mais synthétique et claire, l'état de l'art et les actions nécessaires et souhaitées en matière d'utilisation des TIC en faveur de la diversité linguistique, *Net.lang* a pour ambition de traiter des aspects, tant au niveau du contenu

que du contenant, relatifs à la présence de toutes les langues de la planète dans les technologies de l'information et de la communication, et notamment sur internet. Il devrait ainsi montrer les enjeux du multilinguisme dans le cyberspace.

La capacité à analyser, mesurer, informer sur l'état des langues dans le cyberspace afin d'aider à la définition de politiques publiques de défense, de revitalisation et de partage culturel est un souci constant des institutions qui ont accompagné le Réseau Maaya dans la réalisation de cet ouvrage. Dans le cadre de l'étude sur le multilinguisme dans le cyberspace de 2007 que l'Unesco avait commandé à l'Union Latine, cette dernière institution avait demandé à Marcel Diki-Kidiri l'écriture d'une brochure spécifique dans laquelle on trouverait des pistes concrètes pour doter chaque langue de tous les éléments nécessaires à sa présence sur l'internet, même pour celles peu outillées ou celles dites « orales ». C'est ainsi que le livre « *Comment assurer la présence d'une langue dans le cyberspace* »¹ est devenu en peu de temps un ouvrage de référence synthétisant en quelques pages les étapes essentielles pour permettre à toute langue d'emprunter le chemin numérique.

Avec l'évolution très rapide des technologies comme des enjeux géopolitiques et culturels du réseau, il fallait trouver un successeur à ce travail qui prenne en compte autant l'analyse que les perspectives. C'est ainsi que le Forum International de Bamako sur le Multilinguisme qui s'est tenu en février 2009 à l'initiative de l'Académie Africaine des Langues et du Réseau international pour la diversité linguistique Maaya dans le cadre de l'Année internationale des langues, a recommandé l'édition d'un « *manuel didactique en plusieurs langues montrant à un public non averti les enjeux liés à la présence des langues dans le cyberspace* ».

Un an plus tard, Maaya réunissait à Paris une vingtaine d'experts parmi les plus connus, grâce au concours de l'Unesco et sous la coordination d'un premier Comité scientifique², dans le but de définir les thèmes à traiter et les experts pouvant les développer. Ces deux jours de séminaire ont tracé le cadre, les objectifs, listé les experts en essayant de les prendre

1 <http://unesdoc.unesco.org/images/0014/001497/149786f.pdf>

2 Constitué alors par Daniel Prado (Union latine), Louis Pouzin (EuroInc), Marcel Diki-Kidiri (Association YSB Sângô), Daniel Pimienta (Funredes) et Yoshiki Mikami (Language Observatory).

dans le monde entier et d'assurer le meilleur équilibre possible entre les genres, les âges et les approches techniques, culturelles ou politiques.

Initialement prévu pour février 2011, l'ouvrage connaissait un certain nombre de retards faute d'une structure pouvant assurer un suivi adéquat. Les experts sont souvent occupés par de multiples sollicitations, et la rédaction d'un livre de cette ampleur demande l'apport de connaissances éditoriales spécifiques, ainsi qu'un suivi régulier des auteur(e)s, des traducteurs et traductrices, sans parler des compétences techniques de la typographie et de l'élaboration de documents numériques. C'est ainsi que l'Union Latine a décidé d'apporter un soutien permettant de confier le travail à *C&F éditions*. Nicolas Taffin, Hervé Le Crosnier et Laurent Vannini, les coordonateurs de cet ouvrage, assistés de John Rosbottom et Laura Kraftowitz pour l'édition des textes anglais et les traductions vers cette langue, et de Kathleen Ponsard pour la maquette, ont relevé le défi. Ils lui ont non seulement donné corps, mais aussi le titre, l'image, l'épaisseur et la rigueur qui feront de *Net.lang* un ouvrage majeur en matière de politiques linguistiques.

À la suite des premiers appuis institutionnels apportés par l'Unesco et l'Union latine qui ont permis la genèse de *Net.lang*, ce livre a bénéficié également de l'enthousiasme de l'ANLoc (*African Network for Localisation* - Réseau africain pour la localisation), en partenariat avec le CRDI (Centre de recherches pour le développement international) ; et de l'Organisation internationale de la Francophonie.

Le séminaire de lancement du projet nous a également permis de réfléchir à ce qu'est un livre collectif. S'agit-il uniquement d'un objet imprimé, ou bien d'une collection de textes qui demandent à vivre, se répandre et trouver eux-mêmes leur route dans le cyberspace ? Si l'imprimé permet de trouver place dans les bibliothèques, d'assurer le recensement bibliographique, et de donner statut et insertion aux textes ici réunis, c'est la diffusion numérique qui nous est apparue comme l'élément essentiel. Éditer un livre, c'est avant tout réunir des auteurs, définir un sommaire cohérent, assurer la lisibilité des textes et susciter l'intérêt du lecteur. Peu importe ensuite le support. Ce livre est donc disponible sous de multiples formats numériques et un format imprimé.

Mais un second problème se posait à nous : à partir de combien de langues un livre est-il réellement « multilingue » ? Ou plutôt, comment la

plasticité des versions numériques nous permet-elle de commencer par deux langues véhiculaires... tout en ouvrant la porte à des traductions totales ou partielles dans toutes les langues et par toutes les structures intéressées. Nous avons choisi le français, langue des éditeurs, et l'anglais... pour des raisons que tout le monde comprendra. Mais notre véritable objectif est de voir paraître aussi vite que possible des versions dans de multiples langues.

Pour cela nous nous appuyons sur les dispositions légales des licences Creative Commons by-sa, qui autorisent les travaux dérivés (donc les traductions) dès lors que l'ouvrage et les auteurs des articles sont mentionnés, et que l'édition ainsi réalisée est elle aussi offerte en partage sous la même licence. Ceci permet des travaux éditoriaux mis en vente, mais sans pouvoir bloquer la circulation des textes dans toutes les langues. Nous remercions d'ores et déjà Evgeny Kuzmin qui prépare une édition en russe avant même que les versions originales ne soient publiées. Nous espérons que d'autres acteurs, à titre privé ou au nom d'institutions, effectueront le même travail pour d'autres langues. Le site web qui est mis en place conjointement à la parution de *Net.lang* (<http://net-lang.net>) permet la traduction et la publication de chaque article isolément, et donc d'organiser la collaboration pour obtenir une traduction complète, qui peut alors être transformée en livre numérique ou imprimé dans la langue et dans le pays concerné.

Peut-être même lisez-vous actuellement cette histoire du projet en castillan, en chinois, en malais, et nous l'aimerions beaucoup en wolof, en quechua ou en tagalog. En tout cas, nous l'espérons, et nous avons pris les précautions techniques et juridiques pour que cela soit possible et repose sur l'initiative propre des lecteurs désirant faire connaître autour d'eux et dans leur propre langue les travaux réunis ici.

Une fois traduits dans de multiples langues, nous espérons surtout que *Net.lang* soit un outil qui puisse aider et soutenir tout responsable de politiques d'aménagement linguistique désireux de doter sa langue des outils lui permettant d'être pleinement présente dans le monde numérique. Que les graines que nous semons avec ce livre fassent réellement germer un cyberspace multilingue ouvert à toutes les langues, tous les peuples et tous les savoirs du monde.

**QUAND LES
TECHNOLOGIES
RENCONTRENT
LE MULTI-
LINGUISME**

PARTIE 1

Le cyberspace présente un nombre incalculable de continents invisibles où des milliards de personnes développent ou prolongent des conversations, des relations, des réseaux, des créations, des traductions, des connaissances, des liens sociaux. Quelles passerelles linguistiques peuvent exister pour faciliter ces liens humains en deçà des contraintes techniques et au-delà des barrières culturelles ?

PRÉSENCE DES LANGUES DANS LE MONDE RÉEL ET LE CYBERESPACE

À peine 5% des langues du monde ont une présence dans le cyber-espace et, parmi ce petit nombre, il existe encore des écarts considérables. Seule une infime poignée de langues privilégiées dispose d'une véritable production de contenus. Cet article prétend mettre en parallèle le poids réel d'une langue (démographie, économie, vitalité, officialité, littérature, traduction, etc.) et sa présence (ou sa capacité à être présente pour celles qui en seraient encore absentes) sur l'internet.

Article original en français.



DANIEL PRADO est l'ancien directeur du département Terminologie et industrie de la langue de l'Union Latine, organisation intergouvernementale de promotion des langues et des cultures latines. Il est aujourd'hui Secrétaire exécutif du Réseau mondial pour la diversité linguistique, Maaya.

DANIEL PRADO

PRÉSENCE
DES LANGUES
DANS LE MONDE
RÉEL ET LE
CYBERESPACE

Selon diverses sources, la diversité linguistique s'éteindrait progressivement. Près de la moitié des langues de la planète pourraient disparaître d'ici la fin du siècle selon l'Unesco [LANGUES 2006]. Claude Hagège [HAGÈGE 2000] nous rappelle qu'à l'heure actuelle une langue disparaît en moyenne toutes les deux semaines, un peu moins selon Calvet [CALVET 2002], qui estime qu'il en meurt dix par an. Ce processus est-il inévitable ?

Nombre de locuteurs

La diminution du nombre des locuteurs étant la principale cause de la disparition d'une langue, il est important de préciser que 50 % des langues de la planète seraient aujourd'hui parlées par moins de dix mille habitants [CRYSTAL 2002]. Bien entendu, cette diminution n'est pas un signe de fatalité en soi. Il arrive que des langues peu parlées connaissent un regain de vitalité tel l'hébreu, aujourd'hui langue d'État et considérée morte pendant des siècles, ou bien l'aïnou, enseignée à l'heure actuelle, et « *qui ne comptait plus que huit locuteurs sur l'île d'Hokkaido à la fin des années 1980* » [DIVERSITÉ LINGUISTIQUE 2005].

Volonté politique

La volonté politique des locuteurs ou du moins des représentants d'une langue peut rendre à celle-ci la vitalité que des facteurs endogènes ou exogènes auraient endormie pendant un certain temps (hébreu, aïnou, catalan, basque, français au Québec, etc.). Les langues qui ont su reprendre une place dans la société et évoluer en matière d'utilisation tant quantitative que qualitative sont, pour la plupart, celles qui se sont dotées d'organismes

de tutelle (publics ou privés). Moins de 3 % des langues bénéficient de la protection d'une entité publique et à peine une centaine ont un statut d'officialité (*de facto* ou *de jure*) dans un pays ou région [LECLERC 2011]. Par conséquent, la survie des langues peu utilisées et dénuées d'une telle protection peut sembler préoccupante.

Contexte socio-économique

N'oublions pas que si la diversité linguistique de la planète est loin d'être homogène – les 74 premières langues de l'humanité sont parlées par 94 % des habitants de la planète [LECLERC 2010] – elle concerne surtout les endroits les moins favorisés avec 70 % des langues qui seraient concentrées dans vingt nations [L'ÉDUCATION 2003], pour la plupart les moins riches et donc les moins susceptibles de pouvoir les soutenir. L'entreprise *Globalization Group* (2010) suggère que 90 % du PIB mondial serait produit par les locuteurs utilisant moins de quinze langues¹.

Bien entendu, ces statistiques ne tiendraient compte que des langues officielles des États recensés, ce qui est pour le moins réducteur, et malgré la prudence affichée par les auteurs quant à leur interprétation nous profitons de cette incise pour souligner l'extrême pauvreté d'indicateurs fiables sur la diversité linguistique – phénomène que nous constaterons également plus tard avec le nombre d'internautes par langue.

Langue écrite, langue orale

À l'heure où l'éradication de l'analphabétisme fait partie des *Objectifs du Millénaire pour le développement* [OBJECTIFS 2005] et dans une société contemporaine largement dominée par l'écrit, il est urgent de se demander comment éduquer des enfants (ou des adultes) parlant des langues dites « orales » si la plupart des supports pédagogiques se basent sur l'écrit. Or, entre 90 et 95 % des langues de la planète n'ont pas d'écriture.

1 Soit l'anglais, le chinois, le japonais, l'allemand, l'espagnol, le français, l'italien, le russe, le portugais, l'arabe, le néerlandais, le coréen, le turc et le polonais.

Mondialisation, urbanisation et société du savoir

Le phénomène de disparition des langues, provoqué par plusieurs facteurs du passé proche ou moins proche (colonisations, génocides, épidémies, guerres, déplacements de populations, interdictions d'utiliser la langue, etc.), s'amplifie aujourd'hui au fur et à mesure que progressent mondialisation (économique, technologique, sociale, politique, etc.) et urbanisation. La communication joue un rôle extrêmement important dans la relation de force entre deux langues concurrentes et ce phénomène s'intensifie à l'ère de l'information, favorisant les langues les mieux dotées ou les plus « prestigieuses » au détriment des autres.

Selon Carlos Leáñez, « *moins une langue a de la valeur [aux yeux de ses locuteurs], moins elle est utilisée et moins elle est utilisée, plus elle perd de la valeur* » [LEÁÑEZ 2005]. C'est précisément l'utilisation professionnelle, administrative, éducative, juridique, etc. d'une langue qui permet de la préserver, car des locuteurs qui sont obligés de changer de langue selon le contexte tendent progressivement à utiliser la langue qui leur permet le plus large éventail d'expressions. Or, la grande majorité des langues de la planète ne seraient utilisées que dans un contexte affectif et local.

Dans notre Société du savoir, une langue perd de la valeur aux yeux de ses locuteurs s'ils n'arrivent pas à trouver la connaissance à travers elle. Autrement dit, si elle est absente du cyberspace, ses locuteurs risquent de se tourner vers l'utilisation d'autres langues.

Les langues sur la Toile

Malgré une avancée certaine du multilinguisme sur l'Internet depuis les années quatrevingt-dix, seule une petite poignée des langues² jouit d'une présence notable sur la Toile. Une chose est certaine, l'anglais³ reste la langue la plus utilisée sur internet, même si sa présence relative (par rapport aux autres langues) diminue, passant de 75 % en 1998 à 45 % en 2007 [UNION LATINE - FUNREDES 2007] et selon différents croisements

2 Les sources varient (allant jusqu'à 10% des langues selon la brochure *Initiative B@bel* de l'Unesco), mais un peu plus d'une centaine de langues paraîtrait être une moyenne communément acceptée.

3 Ou, devrions-nous plutôt dire « anglais et *globish* », car cette dernière est de plus en plus identifiée comme une variante à part entière ?

d'études (car, comme nous l'avons mentionné, nous ne disposons plus d'indicateurs fiables), à près de 30 % aujourd'hui. Soulignons que nous parlons ici de quantité de contenus et non pas du nombre d'internautes.

Déficit des langues sur la Toile. Le cas des langues africaines

Si les grandes langues de communication connaissent une présence satisfaisante sur la Toile, la plupart ont une représentation très symbolique, avec seulement quelques pages qui leur sont dédiées. En 2003, une étude menée par Marcel Diki-Kidiri nous montrait que sur 1 374 sites africains étudiés, seuls 3,22% utilisaient une langue africaine comme langue de communication [DIKI-KIDIRI 2003]. Le *Language Observatory Project* [LOP 2011] annonçait en 2009 un recul des langues autochtones dans ce continent après une brève période de récupération jusqu'en 2005. Et pour cause, considéré – avec l'Asie – comme l'un des deux continents où la diversité linguistique est la plus importante (environ 2 100 langues selon le site Ethnologue, 2011), l'Afrique a vu pendant longtemps les langues coloniales s'imposer comme langues véhiculaires au détriment des autres, et peine désormais à réintégrer ses langues premières dans l'éducation, la vie professionnelle, etc. En outre, très peu de langues africaines possèdent une écriture, elles ne peuvent donc pas être représentées sur la Toile autrement que via des supports multimédias, comme nous le verrons plus loin⁴.

Moteurs de recherche, réseaux sociaux

Le moteur de recherche le plus utilisé et le plus développé en matière d'outils linguistiques, Google, n'offrait quant à lui la reconnaissance linguistique que pour une cinquantaine de langues au mois de mars 2011. Alors même que la langue islandaise (240 000 locuteurs) est reconnue depuis longtemps, beaucoup d'autres ayant entre 10 et 200 millions de locuteurs (bengali, javanais, tamoul, malais, haoussa, yoruba, peul, quechua, etc.) ne le sont pas. Le fameux moteur peut reconnaître une trentaine de langues d'origine européenne, mais à peine une langue africaine et aucune langue d'origine américaine ou océanienne.

⁴ Voir dans cet ouvrage : Tunde Adegbola, *Multimédia, langues des signes, langues écrites et orales*.

Yahoo! ne fait pas mieux avec un peu moins de 40 langues reconnues, dont 8 asiatiques, et aucune langue africaine, américaine ou océanienne.

Peu d'outils sur l'internet sont aussi riches en matière de diversité linguistique que Wikipédia (près de 19 millions d'articles dans près de 300 langues). Malgré les tentatives de Twitter, Facebook, YouTube et autres services internet très utilisés, pour se rapprocher des populations en offrant des versions linguistiques localisées, des services de traduction automatique ou encore de sous-titrage, ceux-ci sont encore loin de satisfaire les besoins autres que ceux des locuteurs d'une petite cinquantaine de langues.

Traduction automatique

À propos de la traduction automatique que beaucoup d'analystes considèrent comme la panacée, y compris Graddol [GRADDOL 2007], car elle permet de nous débarrasser du besoin d'une *lingua franca* et que chacun peut s'exprimer dans sa langue, rappelons qu'à peine une soixantaine de langues disposent de tels systèmes. De plus, lorsqu'ils existent, la plupart des systèmes favorisent le couple langue maternelle-anglais, ou entre une petite dizaine de grandes langues de communication (français, chinois, espagnol, allemand, japonais, russe, etc.). Tous les autres couples de langues disposent soit de systèmes rudimentaires, soit utilisent la langue anglaise comme langue pivot. En tout, seul 1 % des langues de la planète disposeraient d'un système de traduction automatique⁵. Les *lingua franca* ont de beaux jours devant elles [PRADO 2010] et un locuteur de bengali devra pour un bon moment faire appel à une troisième langue pour parler à un locuteur de yoruba ou de quechua.

Quoi qu'il en soit, rappelons que les systèmes de traduction les plus performants à l'heure actuelle sont ceux qui disposent de corpus bilingues suffisants (notamment les systèmes dits statistiques, comme *Google Traduction*). La réalité est qu'à peine une cinquantaine de langues disposeraient d'un assez grand nombre de textes traduits si l'on en croit l'*Index Translationum* de l'Unesco⁶.

⁵ Voir dans cet ouvrage : Joseph Mariani, *Les technologies de la langue en soutien au multilinguisme*.

⁶ Voir Les statistiques des langues vers lesquelles on traduit le plus selon l'*Index Translationum* de l'Unesco : <http://databases.unesco.org/xtrans/stat/xTransStat.a?VL1=L&top=50&lg=1>

Les langues des internautes

Une faible productivité est un risque encouru par les langues dans le cyberspace. Dès lors, leurs locuteurs se tournent vers des langues mieux pourvues en contenus, provoquant la boucle « moins de productivité, moins d'audience ; moins d'audience, moins de productivité ».

Des études réalisées sur les grandes langues internationales de communication nous montrent à première vue que la productivité d'une langue sur la Toile serait liée au nombre d'internautes qui l'utilisent et à leur degré d'alphabétisation numérique [PIMIENTA 2007]. Mais nous ne pouvons pas en être aussi certains pour des langues moins utilisées et nous avons même de fortes raisons d'en douter à la lumière de certains faits.

En effet, les principales statistiques concernant la langue des internautes ne prennent en compte que quelques dizaines de langues [INTERNET WORLD STATS 2011] – nombre proche du total des langues reconnues par Google, et toujours aucune d'origine africaine, américaine ou océanienne. Le nombre d'utilisateurs-locuteurs de la plupart des autres langues étant si insignifiant qu'elles ne sont même pas répertoriées par ces statistiques. Il est alors difficile dans ces circonstances d'établir des barèmes de productivité.

En dehors de toute réserve que nous pouvons exprimer sur les chiffres fournis par le site *Internet World Stats* (*ibid.*) sur le nombre d'internautes par langue⁷, il nous permet néanmoins d'apprécier dans le temps l'évolution de la pénétration d'internet par aire géolinguistique. En mars 2011, les locuteurs des dix langues suivantes seraient, dans l'ordre, les plus nombreux à utiliser internet : anglais, chinois, espagnol, japonais, portugais, allemand, arabe, français, russe et coréen. Il est intéressant de constater la progression de l'espagnol (4^e en 2006, 3^e en 2011), du portugais (8^e en 2006, 5^e en 2011), de l'arabe (absent en 2006, 7^e en 2011), du russe (passant de la 10^e place à la 9^e) et tout particulièrement du chinois, cause principale de la réduction relative du nombre d'internautes anglophones (passant de 30 % en 2006 à 25 % en 2011).

7 Les explications fournies par le site ne nous éclairent pas sur la méthode de déduction de la langue utilisée par les internautes et le nombre d'utilisateurs potentiels mentionné pour certaines langues (anglais, arabe, chinois, français, portugais, notamment) nous paraît erroné.

Manque cruel d'indicateurs sur la diversité linguistique et le cyberspace

Les réserves que nous exprimons sur ces chiffres relancent le débat sur le manque d'indicateurs en matière de diversité linguistique sur Internet, et plus particulièrement le manque de finesse dans le détail de ces locuteurs. S'il est vrai que beaucoup d'internautes doivent recourir à une langue dominante parce que leur langue n'est pas suffisamment outillée pour être représentée sur l'internet ou tout simplement parce qu'elle a peu de « valeur » à leurs yeux, le nombre d'utilisateurs d'autres langues est loin d'être négligeable, quoique difficilement quantifiable à l'heure actuelle.

Langues inadaptées au cyberspace ou cyberspace inadapté aux langues ?

En matière d'outillage, nous savons bien que la représentation d'une langue sur internet n'est pas seulement un simple fait culturel ou quantitatif. Il est à la base surtout technique. Paolillo [PAOLILLO 2005] nous rappelle qu'internet est avant tout un instrument conçu pour la langue anglaise. Par extension, les langues à caractères latins et les cultures occidentales ont su plus rapidement que d'autres trouver un lieu d'expression confortable, même si, ne l'oublions pas, les diacritiques propres aux langues européennes autres que l'anglais n'ont toujours pas leur place partout malgré quelques avancées parfois trop médiatisées comme les actions en faveur des noms de domaine acceptant différents alphabets et diacritiques. L'anglais demeure la langue de programmation, de balisage, de codage, de communication entre serveurs et surtout, la langue de l'informatique. Les langages informatiques se basant sur l'anglais, les informaticiens sont obligés de connaître cette langue pour exercer leur profession.

Mais combien de langues trouvent des contraintes bien plus importantes que les langues utilisant des caractères latins avec diacritiques. Des contraintes liées tant à des problèmes techniques de représentation que culturels d'utilisation des médias propres au cyberspace [DIKI-KIDIRI 2007] ?

Similitudes entre la Toile et l'édition classique

On remarquera une similitude entre le monde de l'édition et le monde du Web. Les statistiques montrent que seulement trente langues publient plus de mille volumes par an, parmi lesquelles six produiraient les deux tiers de la littérature mondiale, l'anglais occupant de loin la place prédominante avec 28 % de la production littéraire mondiale [LECLERC 2011].

Les trente premières langues à avoir bénéficié de la reconnaissance linguistique par Google sont *grosso modo* les grandes productrices traditionnelles de littérature. La Toile semble donc calquer la situation que connaît l'édition classique.

Faut-il pour autant conclure que la Toile ne peut s'alimenter que s'il y a une édition qui la précède ?

L'initiative de *Google Livres* – la plus médiatisée de toutes – et celle d'autres opérateurs publics ou privés de numériser les fonds de bibliothèques⁸ pourrait ne servir qu'à réaffirmer le *statu quo* de la diversité linguistique sur la Toile⁹. Pourtant, nous le savons bien, la Toile permet d'ouvrir la porte à des expressions qui n'intéressent pas les circuits d'édition traditionnelle. Après tout, la publication scientifique en d'autres langues que l'anglais a connu un renouveau grâce à la facilité et au faible coût de publication sur la Toile, les éditeurs traditionnels ne souhaitant pas courir le risque d'éditer des articles qui ne concerneraient qu'un nombre minime de lecteurs. Pourtant ce nouveau type d'édition scientifique n'est pas reflété dans les index internationaux qui montrent précisément le processus inverse, à savoir une progression constante de l'anglais¹⁰.

La Toile, même si elle a permis à des minorités absentes de l'édition traditionnelle de s'exprimer, est loin de refléter le prisme de la diversité culturelle en juste proportion, ne serait-ce qu'au niveau du nombre de locuteurs, et rien ne laisse présager que la situation changera dans les années à venir.

8 Voir dans cet ouvrage : Hervé Le Crosnier, *Bibliothèques numériques*.

9 Voir le portail sur les enjeux de la numérisation d'œuvres (*H)ex-Libris*
<http://www.hex-libris.info>

10 Mais ceci est un autre débat qui concerne plutôt les intérêts commerciaux en jeu dans l'édition scientifique.

Mais, il n’y a pas que la Toile...

Aucune étude à l’heure actuelle ne nous donne un aperçu de la place des langues au niveau mondial sur les supports que nous appellerons ici « informels » que sont le courriel, la messagerie instantanée ou *chat*, les forums, les listes de diffusion, les réseaux sociaux numériques, ou encore les blogues. La plupart des sondages, statistiques et autres analyses sont très parcellaires. Nous pouvons cependant apprécier des tendances, grâce à une analyse croisée. Et l’on constate que la production écrite par ces canaux est bien supérieure à la production de pages web – même si elle est souvent éphémère – et que la place des langues moins usitées y est proportionnellement plus importante.

Les blogues

N’oublions pas que malgré la jeunesse du Web 2.0, s’il y avait 255 millions de sites web traditionnels fin 2010, au même moment il y avait 152 millions de blogues et 600 millions de pages Facebook [PINGDOM 2011]. Or, ces espaces se comportent très différemment des pages web traditionnelles.

En 2006, Funredes appliquait sa méthode d’analyse de façon expérimentale aux blogues repérables et constatait :

...que le blogue connaît une logique de productivité différente par rapport au Web traditionnel selon les populations. En effet, les hispanophones, par exemple, produiraient proportionnellement autant de blogues que de pages web¹¹ par rapport aux anglophones, tandis que cette proportion diminuerait à un tiers pour les francophones et lusophones et à un dixième pour les germanophones.

Faut-il conclure à des réticences culturelles de la part des germanophones, francophones et lusophones à utiliser les blogues (sachant que l’on imagine difficilement qu’il puisse s’agir de contraintes techniques ou économiques, la situation socio-économique de ces trois groupes étant inversement proportionnelle à leur présence comparée) ?

La croissance exponentielle de l’univers des blogues (dans lequel les chiffres sont aussi contradictoires qu’éphémères) laisse croire qu’il faudra

11 Même proportion de blogues et de sites web en espagnol par rapport à l’anglais, soit environ 10 pages en espagnol pour 100 en anglais, que ce soit pour les blogues ou pour les pages web.

encore du temps avant que ce média ne se stabilise. Certaines langues se rattraperont probablement comme elles l'ont fait sur la Toile (le minitel français avait ralenti l'émergence d'un internet francophone jusqu'en 2000, par exemple), et l'on verra ces statistiques évoluer.

Cependant, on peut également penser que le blogue est un phénomène culturel (et politique pour certains pays ou régions) et que certaines cultures rechigneront à se l'approprier, tout comme pour le clavardage ou les forums. A contrario, d'autres y trouveront un moyen d'expression bien plus souple qu'avec d'autres médias.

La Toile, de nature plus institutionnelle que le blogue, le *chat* et les réseaux sociaux, sert plutôt à véhiculer des messages ciblant des collectivités larges et visant à être lus par le plus grand nombre. Le message est donc le plus souvent diffusé dans la ou les langue(s) la ou les mieux comprise(s) (en quantité) par les interlocuteurs visés¹².

En revanche, le blogue est plus généralement le fruit d'initiatives individuelles ou locales permettant à une ou plusieurs personnes d'exprimer idées, sentiments, points de vue ou, tout simplement, de se montrer. En général, les créateurs de blogues s'intéresseront plus à exprimer librement un message qu'à la portée réelle, ce qui les incite à utiliser leur propre langue comme moyen d'expression.

Combien de blogues sont réalisés en langues peu usitées? Nous ne le savons pas pour l'instant, mais une simple recherche sur des moteurs de blogues¹³ permet de constater qu'il y a une émergence de ce média en langues très peu usitées sur la Toile.

Le cyberspace interactif

Le courriel, le clavardage, les forums, les listes de discussion restent un lieu de refuge pour les langues minoritaires, tant que tous les utilisateurs se trouvant dans un même espace maîtrisent la langue en cours, et à

12 À moins qu'il ne s'agisse parfois de décisions des dirigeants qui, appartenant à des catégories socioprofessionnelles leur permettant de maîtriser les grandes langues de communication, sont souvent tentés de les utiliser en détriment des autres. Il est fréquent de constater que des organismes internationaux et entreprises multinationales n'utilisent qu'une ou deux grandes langues pour communiquer alors que leur public utilise majoritairement d'autres langues.

13 Par exemple <http://blogsearch.google.fr>

condition qu'un locuteur d'une autre langue n'intervienne pas [PRADO 2005]. En effet, les langues jouissant d'un plus grand « prestige » s'imposent dès que l'un des interlocuteurs ne comprend pas la langue parlée à l'intérieur d'une communauté. Cette réalité qui est connue dans les discussions des forums scientifiques à propos de la prééminence de l'anglais¹⁴, l'est également dans des forums moins formels pour des langues parlées dans des aires où une autre langue est soit langue officielle, soit langue véhiculaire.

Mais les comportements semblent être différents selon les médias. Pour Paolillo [PAOLILLO 2005], certaines populations (les locuteurs du penjabi ou de l'arabe du golfe par exemple) utilisent plus volontiers leur langue maternelle lors des sessions de clavardages que dans la rédaction de courriels, notamment les populations bilingues (parlant donc leur langue maternelle et la langue officielle de leur pays).

Divers forums internationaux et études ont mis en évidence le rôle des médias informels contre ce phénomène tant décrié par les pays émergents par le passé, à savoir la « fuite des cerveaux ». Aujourd'hui, les diasporas conservent le contact avec leur pays d'origine et contribuent à leur développement « en ligne » et dans leur langue.

De nombreux forums ou listes de discussion à caractère régional ou national (voire ethnique ou linguistique) réunissent aussi bien les spécialistes résidant dans leurs pays que ceux qui ont émigré et souhaitent garder des liens. Bien entendu, à condition qu'ils appartiennent à une même communauté linguistique, ces conversations ont lieu non pas en anglais, français ou espagnol, par exemple (ou toute autre langue véhiculaire ou officielle de leurs pays d'origine), mais en penjabi, créole ou guarani, alors que la Toile réserve peu de place à ces langues.

Il n'y a pas que le contenu, il y a aussi le contenant

Internet n'est pas neutre culturellement. Ses formats, sa façon de représenter la réalité, sa topographie, sa gouvernance, ses protocoles et normes,

14 L'anglais finit souvent par s'imposer dans un forum dès qu'un allophone intervient, même si la majorité des locuteurs maîtrisent une autre langue. Ce phénomène est également vérifiable avec de grandes langues de communication rayonnant sur d'autres aires linguistiques (francophonie, lusophonie, russophonie, arabophonie, etc.).

etc. restent liés au milieu anglophone de sa naissance. Internet demeure un espace où la culture anglo-saxonne règne en terrain connu, mais pas seulement du fait de la prééminence de sa langue.

Les formats utilisés, le flux des messages, les façons de combiner le texte, l'image et le son, la taille des écrans, l'utilisation des claviers, la prédominance de l'écrit sur l'oral, etc. sont autant de facteurs qui peuvent ne pas toujours correspondre aux cultures qui souhaitent se l'approprier.

Paolillo (*ibid.*) nous rappelle que les Maoris n'avaient pas accepté les bibliothèques numériques pour des questions exclusivement culturelles, « *la grande disponibilité de l'information [étant] protégée dans la culture maorie (...)* ». Aussi nous pouvons nous poser la question de savoir si la Toile, les forums, les blogs, les listes de diffusion ne vont pas parfois à l'encontre des principes ou valeurs culturelles d'un peuple, et de ce fait, seront moins utilisés par ces cultures, voire pas du tout.

Les formats propres à internet sont loin d'être adaptés aux langues non écrites. Ces langues sont-elles absentes du cyberspace? Dans *Comment assurer la présence d'une langue dans le cyberspace?* Marcel Diki-Kidiri nous montre comment une langue qui n'a pas d'écriture peut emprunter les voies du cyberspace [DIKI-KIDIRI 2007]; mais qu'en est-il de celles dont les locuteurs ne pourront ou ne souhaiteront pas parcourir ces voies?

En dehors de l'écrit

Aujourd'hui, tout type d'information ou de communication peut emprunter les voies électroniques réservées jadis à la seule écriture. La téléphonie sur IP, la radio et la télévision numériques, les téléchargements de contenus audio et vidéo, les sites d'hébergement de vidéo (du type YouTube), la diffusion en flux (*streaming*), etc. font partie du quotidien, du moins dans les pays ou régions disposant d'un parc informatique conséquent et d'un accès à internet facile, peu onéreux et à haut débit. Ainsi, de multiples possibilités s'ouvrent pour la représentation des langues – y compris celles n'ayant pas d'écriture – dans le cyberspace.

N'oublions pas que la téléphonie mobile connaît actuellement un véritable succès en Afrique et que la radio reste l'outil de communication par excellence dans un continent où la presse écrite (notamment en Afrique subsaharienne) est faiblement présente.

L'internet non textuel peut constituer une alternative valable pour les populations ne disposant pas de langues écrites ou possédant des langues mal ou peu reconnues par les systèmes informatiques (problèmes d'encodage, de polices, de claviers, de logiciels, etc.)

Fracture numérique

Encore faut-il que les populations en aient l'accès, puissent produire du matériel audiovisuel et surtout puissent retrouver ces informations. Or, la cartographie mondiale de l'accès au cyberspace montre combien la fracture numérique est liée à la fracture sociale. Internet est certes devenu un outil de la vie quotidienne pour les populations urbaines des pays industrialisés, mais il demeure inconnu pour cinq individus sur sept. En effet, plus de cinq milliards de personnes n'avaient pas accès à Internet fin 2010 [PINGDOM 2011] et la répartition est très inégale : à peine 10 % des Africains sont connectés (sachant que la grande majorité des utilisateurs de ce continent se concentre en Afrique du Sud et sur le bassin méditerranéen) et 25 % des Asiatiques, contre 80 % pour l'Amérique du Nord (Mexique exclu) et 65 % pour l'ensemble de l'Europe.

L'audio et la vidéo

Les solutions à envisager pour réduire cette inégalité d'accès sont plus d'ordre financier, politique et d'alphabétisation numérique que technique. En effet, les services de téléphonie sur IP comme Skype¹⁵ et Messenger¹⁶ (qui acceptent à la fois la voix et la vidéo) sont aujourd'hui à la portée technique de la plupart des internautes, car les contraintes linguistiques sont mineures. De même, la webdiffusion – radio et télévision numériques, baladodiffusion, etc. – est devenue un instrument facilement maîtrisable et moins gourmand en bande passante.

Vœux

La barrière qui empêche 95 % des langues de la planète d'être présentes dans le cyberspace, à savoir l'absence d'écriture ou la non-adaptation

¹⁵ <http://www.skype.com>

¹⁶ <http://messenger.msn.fr>

d'une langue aux TIC, pourrait disparaître. Pour cela, bien sûr, il faut tout d'abord disposer d'ordinateurs adaptés et d'une connexion à haut débit, mais surtout faire en sorte que les populations cibles s'approprient la technique pour en faire des outils utiles à leur développement.

Nul doute que si l'Agenda de Tunis pour la Société de l'information [SMIS 2005] était suivi¹⁷, le danger de disparition des langues reculerait, car en devenant des instruments de communication, elles reprendraient de la valeur. Cependant, accéder aux TIC n'est pas tout, il faut ensuite s'approprier la technique et donc se confronter à un chemin parsemé d'obstacles techniques, culturels ou encore financiers [PIMIENTA, BLANCO 2005].

La fabuleuse diversité linguistique de la planète est en majorité absente du cyberspace. Et pourtant, c'est ce cyberspace qui peut donner une deuxième chance aux langues mourantes... Ou les mettre à mort.

Il est urgent de mettre à plat toutes les informations manquantes par le biais d'indicateurs fiables et complets. Il est urgent de proposer des politiques de sensibilisation auprès des principaux acteurs de langues en danger. Il est urgent de donner à la diversité linguistique les instruments pour sa protection et ceci passe, dans ce XXI^e siècle naissant, en premier lieu par le cyberspace.

BIBLIOGRAPHIE

[CALVET 2002] CALVET, Louis-Jean. *Le marché aux langues*. Paris: Plon, 2002. ISBN 2-259-19660-8.

[CRYSTAL 2002] CRYSTAL, David. *Language Death*. Cambridge: Cambridge University Press, 2002. ISBN 0521012716.

[DIKI-KIDIRI 2003] DIKI KIDIRI, Marcel et BABOYA EDEMA, Atibakwa. Les langues africaines sur la Toile dans Les Cahiers du Rifal, n° 23, *Le traitement automatique des langues africaines*, Bruxelles, Agence intergouvernementale de la Francophonie et Communauté française de Belgique, novembre 2003, ISSN 1015-5716. <http://www.rifal.org/cahiers/rifal23/rifal23.pdf>

[DIKI-KIDIRI 2007] DIKI-KIDIRI, Marcel. *Comment assurer la présence d'une langue dans le cyberspace?* Paris, Unesco, 2007, CI-2007/WS/1 – CLD 31084. <http://unesdoc.unesco.org/images/0014/001497/149786f.pdf>

Ethnologue, languages of the world, 2011. http://www.ethnologue.com/ethno_docs/distribution.asp?by=area

17 Notamment les recommandations proposant un accès universel aux TIC, une réduction des coûts de connexion et l'élargissement des bandes passantes.

Globalization Group inc. *Top Languages by GDP*, 2010. <http://www.globalization-group.com/edge/2010/03/top-languages-by-gdp/>

[GRADDOL 2007] GRADDOL, David. *English next*, Royaume-Uni, 2007. <http://www.britishcouncil.org/learning-research-english-next.pdf>

[HAGÈGE 2000] HAGÈGE, Claude. *Haïte à la mort des Langues*. Paris : Odile Jacob, 2000. ISBN 2-7381-0897-0

[INTERNET WORLD STATS 2011] Internet World Stats. *Internet World Users by Language*, 2011. <http://www.internetworldstats.com/stats7.htm>

[LANGUE 2006] *Langues en danger*, Paris, Unesco, 2006. <http://www.unesco.org/new/fr/unesco/themes/languages-and-multilingualism/endangered-languages/>

[L'ÉDUCATION 2003] *L'éducation dans un monde multilingue : Les contextes multilingues : un défi pour les systèmes éducatifs*, Paris : Unesco, 2003. ED-2003/WS/2. <http://unesdoc.unesco.org/images/0012/001297/129728f.pdf>.

[LEAÑES 2005] LEAÑEZ, Carlos. «Español, francés, portugués: ¿equipamiento o merma?» dans *Congreso internacional sobre lenguas neolatinas en la comunicación especializada*, México, Colegio de México, 2005. ISBN 968-12-1179-0. http://dtil.unilat.org/cong_com_esp/comunicaciones_es/leanez.htm#a

[LECLERC 2011] LECLERC, Jacques. «L'expansion des langues» dans *L'aménagement linguistique dans le monde*, Québec, TLFQ, Université Laval, 24 janvier 2011. http://www.tlfq.ulaval.ca/axl/langues/2vital_expansion.htm

[LECLERC 2010] LECLERC, Jacques. «L'inégalité des langues» dans *L'aménagement linguistique dans le monde*, Québec, TLFQ, Université Laval, 2 mai 2010. http://www.tlfq.ulaval.ca/axl/langues/1div_inegalite.htm

[LOP 2011] LOP, *Language observatory project*, 2011. <http://www.language-observatory.org/>

[OBJECTIFS 2005] *Objectifs du Millénaire pour le développement*, New York, Nations unies, 2005. <http://mdgs.un.org/unsd/mdg/Home.aspx>

[PAOLILLO, PIMIENTA 2005] PAOLILLO, John, PIMIENTA, Daniel et PRADO, Daniel. *Mesurer la diversité linguistique sur Internet*. Paris, Unesco, 2005, CI.2005/WS/06. <http://unesdoc.unesco.org/images/0014/001421/142186f.pdf>

[PAOLILLO 2005] PAOLILLO, John. Diversité linguistique sur Internet : examen des biais linguistiques dans *Mesurer la diversité linguistique sur Internet*. Paris, Unesco, 2005, CI-2005/WS/06 CLD 24822. <http://unesdoc.unesco.org/images/0014/001421/142186f.pdf>

[PIMIENTA, BLANCO 2005] PIMIENTA, Daniel, BLANCO Alvaro. *Le chemin parsemé d'obstacles des technologies de l'information et de communication (TIC) vers les TIC pour le développement humain (DH) et l'approche par processus*, Saint-Domingue, Funredes, 2005. <http://www.funredes.org/presentation/TICpDHf.ppt>

[PIMIENTA 2005] PIMIENTA, Daniel. Diversité linguistique dans le cyberspace : modèles de développement et de mesure dans *Mesurer la diversité linguistique sur Internet*. Paris, Unesco, 2005, CI-2005/WS/06 CLD 24822. <http://unesdoc.unesco.org/images/0014/001421/142186f.pdf>

[PIMIENTA 2007] PIMIENTA, Daniel. *Fracture numérique, fracture sociale, fracture paradigmatique*, Saint-Domingue, Funredes, juillet 2007. http://funredes.org/mistica/francais/cyberotheque/thematique/fracture_paradigmatique.pdf

[PINGDOM 2011] Pingdom. *Internet 2010 in numbers*, janvier 2011. <http://royal.pingdom.com/2011/01/12/internet-2010-in-numbers>

[PRADO 2010] PRADO, Daniel. « La traduction automatisée : le cas des langues romanes » dans *Traduction et mondialisation*, revue Hermès n°56, CNRS éditions, Paris, 2010, ISBN : 978-2-271-06992-4, ISSN : 0767-9513.

[DIVERSITÉ LINGUISTIQUE 2005] PRADO, Daniel. Diversité linguistique dans le cyberspace. Le contexte politique et juridique dans *Mesurer la diversité linguistique sur Internet*. Paris, Unesco, 2005, CI-2005/WS/06 CLD 24822. <http://unesdoc.unesco.org/images/0014/001421/142186f.pdf>

[SMIS 2005] Sommet mondial sur la société de l'information. *Agenda de Tunis pour la société de l'information*, UIT, 18 novembre 2005, WSIS-05/TUNIS/DOC/6(Rév.1)-F. <http://www.itu.int/wsis/docs2/tunis/off/6rev1-fr.html>

[UNION LATINE - FUNREDES 2007] Union Latine-Funredes. *Langues et cultures sur la Toile 2007*, Paris, 2007. http://dtl.unilat.org/LI/2007/index_fr.htm

Vitalité et disparition des langues, Paris, Unesco, 2003. <http://www.unesco.org/culture/ich/doc/src/00120-FR.pdf>

L'ANGLAIS NE SERA PAS LA *LINGUA FRANCA* DE L'INTERNET

Dans les années 1990, l'anglais était tellement dominant sur le web que certains en avaient d'ores et déjà fait la langue véhiculaire incontestée de la mondialisation. Aujourd'hui, la part de l'anglais est tombée en dessous de la barre symbolique des 50%, notamment en raison de la montée en puissance des langues des BRIC (Brésil, Russie, Inde, Chine) et de la dés-occidentalisation progressive du centre de gravité de la planète. La question change donc de nature : que signifie l'émergence du multilinguisme sur cette interface ?

Article original en français.



MICHAËL OUSTINOFF, maître de conférences habilité à diriger des recherches en traductologie à l'Institut du monde anglophone de l'Université Paris 3 Sorbonne Nouvelle, est actuellement en délégation à l'ISCC, l'Institut des sciences de la communication du CNRS. Son troisième ouvrage, *Traduire et communiquer à l'heure de la mondialisation* est paru chez CNRS Éditions en 2011.

MICHAËL OUSTINOFF

L'ANGLAIS
NE SERA PAS
LA LINGUA FRANCA
DE L'INTERNET

Au début des années 1990, au moment où Internet commence à s'étendre à la planète entière, la présence de l'anglais est écrasante sur la Toile. Nombreux sont ceux à y voir alors la confirmation de l'inexorable ascension de l'anglais en tant que « langue planétaire », à commencer par David Crystal dans *English as a Global Language* [CRYSTAL 1997].

Qu'elle le soit devenue, c'est incontestable, mais c'est une toute autre affaire d'en déduire qu'elle est appelée à devenir le vecteur *unique* de la communication internationale à l'heure de la mondialisation : une telle thèse résiste de moins en moins à l'épreuve des faits.

Signe des temps : c'est au sein du monde anglophone lui-même que les critiques se font aujourd'hui les plus argumentées et les plus radicales, à commencer par la *British Academy* dans un rapport récent [BRITISH ACADEMY 2009] et le non moins sérieux *Department of Energy* (DOE) aux États-Unis, qui a lancé en juin 2010 la plateforme multilingue *WorldWideScience.org* en partant de la constatation que la science – en particulier les sciences dites « dures » – ne se fait plus par le seul truchement de l'anglais.

L'évolution des langues présentes sur internet est à cet égard éclairante : l'anglais est aujourd'hui descendu bien en dessous du seuil symbolique des 50 %, et même de celui des 30 % du volume total. Paraphrasant la célèbre formule d'Umberto Eco pour qui la langue de l'Europe, c'est la traduction, il faudrait dire que la *lingua franca* de l'internet, c'est le plurilinguisme, et, par voie de conséquence, la traduction. C'est à une véritable rebabélisation du monde à laquelle nous assistons, et internet ne fait que confirmer cette tendance.

Rien d'étonnant à cela : nous avons désormais suffisamment de recul pour comprendre les facteurs expliquant un tel renversement de paradigme, qui demande à être appréhendé en des termes nouveaux, en mettant en œuvre une approche résolument pluridisciplinaire [OUSTINOFF, NOWICKI, MACHADO DA SILVA 2010] et non uniquement linguistique. Car derrière la question des *lingua franca* se profilent des enjeux à la fois économiques, culturels et géopolitiques majeurs : le cyberspace en est une des manifestations les plus frappantes.

LA RÉSISTIBLE ASCENSION DU TOUT-À-L'ANGLAIS ET LE « PARADOXE DE LA LANGUE DOMINANTE »

La chute de la part relative de l'anglais sur internet est spectaculaire [PAO-LILLO, PIMIENTA, PRADO, ET AL. 2005] : elle se situerait autour de 27,3 % en juin 2010 selon *Internet World Stats*, dont les données sont consultables en ligne¹. Si l'on se tient à ces chiffres, cette diminution tient essentiellement à quatre facteurs. Tout d'abord, au taux de croissance des langues en présence. Sur la période 2000-2010, on s'aperçoit que l'anglais n'a progressé que de 281,2 % (ce qui reste néanmoins considérable) alors que le chinois, désormais la deuxième langue la plus représentée sur internet avec une part de 22,6 %, a connu un taux de progression de... 1277,4 %. Toutefois, une langue n'est pas en mesure d'en dépasser une autre uniquement en raison d'un taux de croissance plus élevé : l'arabe connaît la progression la plus impressionnante sur la même période, soit 2501,2 %, mais n'en est qu'à une part de 3,3 %, ce qui le place en septième position seulement, derrière l'espagnol (numéro trois, avec 7,8 %), le japonais (quatrième avec 5 %), le portugais (cinquième, 4,2 %) et l'allemand (sixième, 3,8 %) et devant le français (huitième, 3 %), le russe (neuvième, également 3 % environ) et le coréen (dixième, 2 %). Les dix premières langues totalisent 82 % de l'ensemble, les autres ne représentant pour l'instant que les 18 % restants.

Le deuxième facteur, c'est le taux de connexion à internet (*Internet Penetration per Language*), qui varie selon les infrastructures et le degré de développement des pays. Au Japon, ce taux est de 78,2 %, ce qui représente plus de 99 millions d'utilisateurs et explique que la langue japonaise

1 Internet World Stats <http://www.internetworldstats.com/stats7.htm>

vienne en quatrième position, avec un taux de croissance relativement modeste (110,6%), alors que l'arabe, avec un nombre de locuteurs potentiels bien plus élevé (347 millions, alors que le japonais n'en compte que 126, pour l'essentiel concentrés sur le territoire du Japon), vient loin derrière, en raison d'un « taux de pénétration » de seulement 18,8%. De ce point de vue, l'arabe est comparable au français. Le nombre de locuteurs est sensiblement le même (347 millions de locuteurs répartis sur l'ensemble de la planète) et, puisque le taux de pénétration est également très proche (17,2%), les deux langues se talonnent actuellement de près.

Le troisième facteur, c'est par conséquent le nombre total des locuteurs d'une langue, qui représente un vivier potentiel destiné à grossir le nombre des utilisateurs d'internet une fois la « fracture numérique » résorbée. Le classement actuel des dix premières langues n'a rien de figé : il est au contraire dynamique et susceptible d'évoluer dans le temps. Il n'est pas besoin d'être grand clerc pour se rendre compte que l'arabe, avec un taux de progression de 2501,2% contre 398,2% pour le français, ne peut que le distancer, à condition cependant que le taux de pénétration augmente. À l'inverse, le français, même au rythme actuel, pourrait lui passer devant, si le taux de pénétration, notamment en Afrique, venait à croître plus vite que celui de l'arabe, et ainsi de suite. À l'inverse, plus le taux de pénétration est déjà élevé, plus il y a de chances de voir les langues en question rétrograder dans le classement : c'est le cas du japonais et de l'allemand, par exemple, avec 78,2% et 78,6% respectivement. Ce sont également des langues qui obtiennent les meilleurs scores en raison du statut de grandes puissances économiques que sont le Japon et l'Allemagne, ou du niveau de vie des autres pays germanophones (Autriche, Suisse, etc.).

En combinant ces trois facteurs à un quatrième, celui du développement économique et de la puissance géopolitique qui va de pair, on voit clairement que le modèle du tout-à-l'anglais, naguère dominant, se verra dans un proche avenir concurrencer en haut de la pyramide par le chinois, l'espagnol, l'arabe, le hindi ou l'indonésien, ces deux dernières langues n'apparaissant pas encore au palmarès.

Mais il y a un cinquième facteur à prendre en compte, essentiel, et que *Internet World Stats* laisse délibérément de côté : « *Si de nombreuses personnes sont bilingues, ou multilingues, nous avons choisi de n'assigner qu'une seule langue par individu, de façon à ce que la somme des locuteurs*

par langue soit égale à la population mondiale»². Il ne faut en effet pas oublier que, puisqu'il existe entre 6000 à 7000 langues dans le monde pour seulement environ deux cents États, le monolinguisme n'est pas la règle, mais l'exception. À quoi l'on ajoutera, en tant que cas particulier nullement négligeable, celui de l'intercompréhension : un lusophone aura directement accès à 4,2 % de l'ensemble d'internet, ce qui peut sembler peu, mais également à la partie hispanophone (7,8 %), car il est très facile de passer d'une de ces langues à l'autre, notamment à l'écrit. C'est donc à près de 12 % de l'ensemble qu'il pourra se connecter, ce qui est déjà considérable : c'est autant que le russe, le français, l'arabe, et l'allemand réunis.

Supposons que le lusophone en question soit un Brésilien, qu'il ait fait ses études en France et que, de surcroît, il maîtrise l'anglais, cas relativement répandu au Brésil. C'est non pas à 4,2 % de l'ensemble qu'il aura accès (portugais seul), ni à 12 % (portugais + espagnol), ni même à 15 % (en y ajoutant le français) mais bien à 42 % (portugais + espagnol + français + anglais). Prenons son homologue chinois : il aura accès à près de 50 % de l'ensemble (chinois + anglais). Mais ce ne sont là que des données quantitatives. Imaginons un internaute francophone, mais maîtrisant également, en sus de l'anglais, l'allemand : il n'aura accès qu'à une part quantitativement plus modeste (37 %), mais ses trois langues sont trois langues majeures de l'Union européenne. Autrement dit, la dimension *qualitative* est également à prendre en compte.

À l'inverse, les anglophones dont l'anglais est la première langue ont tendance à être massivement monolingues, en raison même du modèle du tout-à-l'anglais encore aujourd'hui présenté comme la panacée universelle. Ils n'auront par conséquent accès qu'à 27 % de l'ensemble. La conclusion s'impose d'elle-même : dans un monde désormais multipolaire, où la mondialisation actuelle s'est accompagnée d'un développement sans précédent des technologies de l'information et de la communication (TIC), ne maîtriser que la *lingua franca* c'est être sous-informé, phénomène que Louis-Jean Calvet a appelé le « *paradoxe de la langue dominante* » [CALVET 2007]. Or, dans le monde d'aujourd'hui comme dans celui d'hier, être sous-informé constitue un handicap. Dans le cas particulier d'internet, lorsqu'au début l'anglais représentait plus de 80 % de l'ensemble, cette sous-information pouvait être considérée comme marginale : désormais, elle ne l'est plus. La donne est radicalement différente.

² <http://www.internetworldstats.com/stats7.htm>

ÉTATS-UNIS, GRANDE BRETAGNE, MÊME COMBAT : NON AU TOUT-À-L'ANGLAIS

À l'heure de la mondialisation, la sous-information est un luxe que l'on ne peut plus se permettre. Voilà pourquoi la maîtrise de la *lingua franca* dominante est devenue insuffisante. À cet égard, il est assez piquant de constater que l'on continue de promouvoir le tout-à-l'anglais dans un pays comme la France – bien d'autres pourraient être cités – au moment même où le monde anglophone commence à remettre fondamentalement ce modèle en question [MARTEL 2010].

On retiendra trois moments clés de cette remise en cause. On fera remonter le premier à l'étude magistrale faite pour le compte du British Council par David Graddol sous le titre *The Future of English? The Popularity of the English Language in the 21st Century* [GRADDOL 1997]. À notre connaissance, c'est la première analyse de fond réalisée par un anglophone envisageant que l'avenir de l'anglais en tant que langue internationale « planétaire » (*Global English*) est loin d'être assuré. En effet, rien n'empêche de penser que d'autres langues viendront la concurrencer sous l'effet de la désoccidentalisation du monde et la montée en puissance des économies émergentes comme les BRIC (Brésil, Russie, Inde, Chine). En 1999, ce n'était qu'une hypothèse : en 2011, elle se confirme à l'échelle d'internet. Mais le point fondamental était surtout de mettre en garde les anglophones contre les dangers du *monolinguisme* : victime de son succès, l'anglais s'est étendu à ce point à l'échelle planétaire que le nombre de « locuteurs natifs » a été dépassé par le nombre de locuteurs ayant l'anglais pour deuxième langue. Conséquence : dans un monde devenu multipolaire, la maîtrise de l'anglais devient une simple compétence de base. Ce qui fait la différence, ce n'est donc plus la maîtrise de deux langues (anglais + 1), ce qui devient la norme pour les non-anglophones, mais bien d'*au moins* deux langues (anglais + 2 ou davantage).

Deuxième coup de butoir porté par David Graddol à une autre idée reçue : l'anglais n'est pas destiné à être la *lingua franca* partout dans le monde. Pourquoi utiliser l'anglais comme langue de communication en Amérique latine alors que l'interface espagnol/portugais est la solution, de loin, la plus pratique en raison du fort degré d'intercompréhension entre les deux langues ? C'est souligner l'importance stratégique des *lingua franca* locales. Mais David Graddol va plus loin, en constatant qu'il vaut toujours mieux s'adresser aux autres dans leurs langues maternelles respectives,

notamment en matière commerciale [BEL HABIB 2011]. Si l'on désire faire des affaires en Asie du Sud-Est, l'anglais est certes utile, mais le chinois est désormais une langue de communication majeure dans la région : plutôt que de passer par l'anglais, c'est cette langue « locale » qui servira de *lingua franca*. Qui plus est, les *lingua franca* locales s'internationalisent, notamment dans les relations Sud-Sud : les Chinois se mettent à apprendre le portugais pour commercer avec le Brésil et, réciproquement, les Brésiliens se mettent au chinois. La multiplication des Instituts Confucius pour le chinois, des Instituts Camoëns, pour le portugais, des Instituts Cervantès, pour l'espagnol, etc. s'expliquent à partir de là : le *soft power* n'est plus réservé à la seule langue anglaise, ce que la Chine, notamment, a bien compris. Voilà qui explique, par exemple, l'augmentation spectaculaire de l'enseignement du chinois en France dans l'enseignement public, bien que cette langue ait la réputation d'être bien plus difficile que l'allemand, dont l'enseignement dans le même temps est en déclin.

Troisième et dernier coup de butoir, David Graddol n'exclut pas que l'anglais soit à la merci d'un « scénario catastrophe » à un moment où la défense de la diversité des langues et des cultures est de plus en plus clairement perçue comme une nécessité par l'opinion publique mondiale : « Ces tendances rendent possible un « scénario catastrophe » dans lequel le monde entier se retourne contre l'anglais, en l'associant à l'industrialisation, à la destruction des cultures, aux atteintes aux Droits de l'Homme, à l'impérialisme culturel mondialisé et à l'accroissement des inégalités » [GRADDOL 2000 : 62]. C'est pourquoi la citation suivante de Leonard Orban, alors Commissaire européen au multilinguisme, ne saurait apparaître comme irréaliste mais au contraire d'actualité : « Les salariés devraient donc maîtriser, pour le plus grand bonheur de leurs employeurs, au moins trois langues : celle de leur pays d'origine, l'anglais bien sûr, et une troisième parmi les plus parlées de l'UE – l'allemand, le français, l'espagnol ou l'italien. Sans négliger le russe, l'arabe ou le chinois » [RICARD 2007]. Posé en termes économiques, ce n'est plus le tout-à-l'anglais qui constitue un atout en monnaie sonnante et trébuchante (dont les premiers bénéficiaires étaient jusqu'ici les pays anglophones) mais bien le plurilinguisme, véritable « avantage concurrentiel » de la mondialisation en cours et à venir.

Le deuxième moment clé de cette prise de conscience, c'est la publication en 2009 du rapport *Language Matters* par la *British Academy*, qui vient, pour ainsi dire, enfoncer le clou, cette fois-ci dans le domaine de

la recherche, en s'alarmant du niveau de plus en plus préoccupant des chercheurs britanniques en matière de langues étrangères (p. 3) :

Dans les Humanités, par exemple, les recherches dans des domaines comme l'histoire et la philosophie doivent s'appuyer sur des études effectuées dans d'autres langues, qui n'ont pas été traduites ou sont peu susceptibles de l'être. Dans les sciences sociales, les études comparées et les travaux transnationaux en sciences politiques, en sociologie ou en économie du développement exigent des compétences en d'autres langues. Les chercheurs de toutes les disciplines – y compris les sciences de la nature – ont besoin de maîtriser des langues étrangères non seulement à l'écrit mais également à l'oral afin de saisir les occasions d'aller faire des recherches et de travailler à l'étranger ou de collaborer avec des partenaires à l'échelle internationale. En raison de la multiplication des travaux réalisés en collaboration, ainsi que de l'importance considérable des fonds leur étant aujourd'hui consacrés par les agences aussi bien nationales qu'internationales, un manque de connaissances en matière de langues inflige un handicap certain aux chercheurs dans de nombreux secteurs du système universitaire britannique, ce qui contribue à affaiblir la compétitivité du système dans son ensemble.

Qui aurait pu prévoir, il y a de cela seulement trente ans, que de telles déclarations émaneraient un jour d'une institution anglophone aussi prestigieuse que la *British Academy*? Rappelons-nous : le tout-à-l'anglais était censé permettre de se passer des autres langues, à l'instar de la *koinè* dans le monde hellénistique, du latin au Moyen Âge, ou du français à l'époque des Lumières, et bien davantage encore, puisque ces *lingua franca* étaient l'apanage des élites alors que l'anglais serait accessible au plus grand nombre, à l'ère de l'enseignement de masse.

Mais il y a un troisième et dernier moment clé à retenir, qui mériterait un développement à part : c'est celui du lancement, sous l'égide du *Department of Energy* des États-Unis de la plateforme internationale multilingue *WorldWideScience.org*, car ce projet remet en cause le fondement ultime du tout-à-l'anglais, et, plus généralement, le recours à une *lingua franca* unique en matière de communication scientifique à l'échelle planétaire.

LINGUA FRANCA, INFORMATION ET COMMUNICATION À L'HEURE DE LA MONDIALISATION

Avant d'être remis en cause, le tout-à-l'anglais était présenté, jusqu'à la fin des années 1990, comme un modèle dont l'utilité, pour ne pas dire la nécessité, semblaient évidentes. Une célèbre grammaire de la langue anglaise ne craignait d'ailleurs pas d'avancer tout de go : « *English is the world's most important language* » [QUIRCK, AL. 1980 : 2]. Quatre principaux avantages de ce modèle étaient retenus, outre le fait que l'anglais soit la langue internationale de plus grande diffusion :

- l'anglais est la solution la plus pratique, car c'est une « langue facile » ;
- le tout-à-l'anglais est la solution la plus économique ;
- c'est la solution la plus démocratique et la plus équitable ;
- l'anglais, en tant que *lingua franca*, est « culturellement neutre ».

Ce dernier point est exposé de la manière suivante (*ibid.* : 6) :

L'anglais est d'évidence la plus internationale des langues. Même si la mention de cette langue fait immédiatement penser à l'Angleterre d'une part et aux États-Unis de l'autre, elle porte en elle moins de présupposés politiques ou de spécificité culturelle que les autres langues vivantes (notamment le français et l'espagnol).

Autrement dit, non seulement on présuppose ainsi que toutes les langues sont interchangeable, mais également que l'anglais est en quelque sorte la plus interchangeable de toutes, en raison de sa « neutralité ». Voilà ce qui, en définitive, la légitime en tant que *lingua franca*. Néanmoins, ce terme est, en lui-même, ambivalent, si l'on se reporte à la définition qu'en donne le Petit Robert 2011. Au sens originel, la *lingua franca* est en réalité un sabir particulier, désignant une « langue parlée jusqu'au XIX^e siècle dans les ports méditerranéens ; c'est une langue composite à base d'italien central mais aussi de français et d'espagnol, mêlée d'éléments grecs et arabes » dont l'âge d'or se situe aux XII^e et XIII^e siècles. C'est une langue purement utilitaire. À l'inverse, le terme recouvre aujourd'hui, par extension, une tout autre réalité : « Langue véhiculaire utilisée sur une aire géographique assez vaste. Le swahili, lingua franca en Afrique de l'Est. » (*ibid.*) Par « langue véhiculaire » il faut entendre une langue « servant aux communications entre groupes de langue maternelle différente » qui s'oppose à la « langue vernaculaire » (*ibid.*). Une dissymétrie flagrante de l'anglais en tant que *lingua*

franca apparaît immédiatement : contrairement au latin, qui n'était la langue maternelle de personne au Moyen Âge, l'anglais est à la fois langue véhiculaire et langue vernaculaire pour les locuteurs natifs anglophones.

C'est en prenant en compte cette dissymétrie fondamentale que l'on peut s'expliquer pourquoi le cyberspace tend actuellement à un plurilinguisme de plus en plus étendu et non l'inverse, et encore moins au tout-à-l'anglais (anglais + langue vernaculaire), aboutissant au bilinguisme ou au monolinguisme selon que l'on ait, ou non, l'anglais pour première langue. Autrement dit, le cyberspace tend davantage à se ranger du côté de Wikipédia (où l'on dénombre au total 280 langues) que du côté des portails des grandes institutions internationales dont la plus multilingue est l'Union européenne, avec actuellement 23 langues officielles pour 27 pays membres.

Aux antipodes de la langue « hypercentrale » qu'est l'anglais, pour reprendre la terminologie de Louis-Jean Calvet et des langues « centrales » de grande diffusion que sont l'espagnol, le français ou l'arabe, on trouve, par exemple, des articles en navajo sur Wikipédia : il n'y a que 170 717 locuteurs de cette langue [SHIN, KOMINSKI 2010], au demeurant la langue amérindienne la plus parlée aux États-Unis si l'on en croit le dernier recensement officiel de ce pays en 2007. Sauf coup de théâtre retentissant, cette langue a peu de chances de figurer au palmarès des *lingua franca* majeures du XXI^e siècle. Pourtant, cette langue vient en 173^e position, dans la catégorie des langues comptant plus de 1000 articles sur l'encyclopédie collaborative en ligne (très exactement : 2154). Et, naturellement, il n'est pas indifférent de s'exprimer sur internet en navajo plutôt qu'en anglais aux États-Unis, point sur lequel il est inutile de s'attarder sinon pour dire que c'est là une preuve de plus de l'importance de la langue maternelle, en ce sens jamais « culturellement neutre » – y compris, à l'autre bout du spectre, pour un locuteur natif anglophone.

Il est cependant un domaine qui semble pouvoir court-circuiter toutes les différences qui séparent les langues des autres, autrement dit de transcender les « visions du monde » (*Weltanschauungen*) que toutes les langues, selon Wilhelm von Humboldt, contiennent en propre, chacune étant en cela irréductible aux autres. C'est celui de la science, comme l'affirme Descartes au début du *Discours de la méthode* : « Ceux qui ont le raisonnement le plus fort, et qui digèrent le mieux leurs pensées, afin de les rendre claires et intelligibles, peuvent toujours le mieux persuader ce qu'ils proposent,

encore qu'ils ne parlissent que bas-breton, et qu'ils n'eussent jamais appris de rhétorique » [CASSIN 2004 : 466]. Puisque les langues sont à cet égard interchangeables, il justifie ainsi le fait qu'il écrive le *Discours* non pas en latin, la langue dominante de l'époque, mais en français, afin de toucher le plus grand nombre ; et s'il ne s'exprime pas en bas-breton, ce n'est pas parce que cette langue est moins apte à exprimer une pensée complexe au même titre que le latin ou le français, mais bien parce que, ce faisant, il réduirait son lectorat d'autant.

Transposons le même raisonnement dans le monde d'aujourd'hui : si la science est indifférente à la langue dans laquelle on l'exprime, alors autant recourir à la langue la plus diffusée dans le monde, autrement dit, l'anglais. Non que les autres, de la plus centrale (comme l'espagnol ou l'arabe) à la plus périphérique (disons, le navajo) en passant par celles qui comptent le plus de locuteurs (le chinois, l'hindi ou l'indonésien), soient moins dignes de la transmettre : mais aucune n'est mieux en mesure de parvenir au plus grand nombre. Certes, la relation est asymétrique selon qu'on est locuteur natif ou non : pour paraphraser Winston Churchill au sujet de la démocratie, ce n'est là, après tout, qu'un moindre mal.

Le raisonnement semble sans faille : la langue de la science – en particulier les sciences dites « dures » – n'est-elle pas l'anglais ? C'est là, découvre-t-on désormais, une illusion d'optique.

Les États-Unis sont en train de le comprendre. Car ils se rendent compte que leur production scientifique, autrefois loin devant les autres à la première place, est en train d'être rattrapée, vitesse grand V, par celle d'autres pays, à commencer par l'Europe ou les BRIC (Brésil, Russie, Inde, Chine). Et ils ont également compris que la science ne se fait plus seulement en anglais, mais dans d'autres langues.

C'est pourquoi le *Department of Energy* aux États-Unis emboîte le pas à la *British Academy*, sur le versant de la traduction couplée aux nouvelles technologies, en créant, fin juin 2010, la plateforme *Multilingual WorldWideScience.org*, une plateforme multilingue où l'on peut faire des recherches dans 70 banques de données scientifiques provenant de 66 pays à travers le monde dans les langues les plus diverses. Le tout est couplé à un moteur de recherche multilingue et un programme de traduction automatique en neuf langues, à savoir : l'allemand, l'anglais,

le chinois, le coréen, l'espagnol, le portugais et le russe, d'autres langues devant progressivement s'ajouter à cette liste.

Si l'on prend l'exemple de la Chine, voilà ce qu'on trouve expliqué sur le site: «*En 2008, les chercheurs chinois ont publié 110 000 articles dans les revues internationales du Science Citation Index. La même année, ils ont fait paraître 470 000 articles dans des revues nationales en chinois. À défaut d'accéder à ces 470 000 articles, on peut difficilement se faire une idée réaliste concernant l'état des sciences et des techniques en Chine. Cette situation renforce la nécessité de traduction mutuelle entre l'anglais et le chinois et rend plus urgent encore la recherche documentaire multilingue.*»

On ne saurait être plus clair: pour communiquer, il ne suffit certes pas d'être informé [WOLTON 2003], mais encore faut-il, au préalable, avoir accès à l'information tout court. Et dans un monde de plus en plus multipolaire et multilingue n'ayant plus pour centre de gravité l'Occident (même en incluant le Japon dans la fameuse Triade), une *lingua franca*, pour universelle qu'elle soit, ne saurait, à elle seule, remplacer toutes les autres. Ce paradigme, hérité de l'après Seconde guerre mondiale qui a vu l'émergence de l'hyperpuissance américaine, a vécu: les États-Unis eux-mêmes commencent à en prendre acte.

CONCLUSION

On l'aura compris, la question de la *lingua franca* est à replacer dans le cadre plus large de la géopolitique, de même que la traduction, question tout aussi centrale, et dont l'étymologie s'avère éclairante. Le terme *translatio* (latin « transfert ») s'entendait au Moyen Âge, au sens de « traduction », sens que l'anglais « *translation* » a conservé, mais il servait également [CASSIN 2004: 1312] à désigner le transfert du savoir (*translatio studii*) ou du pouvoir (*translatio imperii*). Le savoir comme le pouvoir avaient ainsi été transmis par la Grèce à Rome, puis à l'Occident, phénomène qui se poursuit aujourd'hui en direction d'autres parties du monde, en particulier de l'Asie (Inde, Chine) ou des Sud (notamment l'Amérique latine, les pays Arabes, demain peut-être l'Afrique). Qui dit émergence de nouvelles puissances, dit également redistribution des cartes en haut de la pyramide. C'est ce qui explique que le chinois est en passe de devenir une langue internationale, alors qu'auparavant on ne la parlait guère en dehors de la Chine ou de Taïwan et au sein de la diaspora chinoise.

Le cas du portugais est, à cet égard, révélateur. Comme l'anglais, l'espagnol, l'arabe ou le russe, c'est une *lingua franca* à l'échelle d'une aire linguistique, celle de la lusophonie. Mais c'est aussi l'une des langues majeures des BRIC : en 1960, le Brésil était un pays en voie de développement, comptant environ 70 millions d'habitants. Aujourd'hui, ce pays est la huitième puissance économique mondiale et avoisine les 200 millions d'habitants. De périphérique, le portugais occupe dès lors une place de plus en plus centrale. Dans un monde multipolaire, le maintien de l'anglais comme seule *lingua franca* de référence apparaît de moins en moins évidente.

En 2005, Michael J. Barany [BARANY 2005], mathématicien en poste à Princeton, faisait paraître dans *Business Week* un article intitulé *Le problème de la langue dans les sciences* (*Science's Language Problem*) qui résume très bien ce qui précède :

Demain, alors que les chercheurs parlant couramment l'anglais verront sans nul doute leur nombre décliner dans les laboratoires du monde entier, la véritable camisole de force que représente l'anglais va devenir de plus en plus inconfortable à supporter, au moment même où nous assistons à une explosion du volume des publications scientifiques.

En Inde ou en Chine des infrastructures scientifiques de niveau international voient le jour, et de plus en plus de découvertes ne seront diffusées que dans leurs propres langues. Ce qui conduira à leur méconnaissance ou leur sous-évaluation partout ailleurs, tout comme les travaux publiés actuellement en japonais ou en français parviennent rarement à impressionner les chercheurs des États-Unis qui ne maîtrisant pas ces langues. Quant aux articles qui ont subi plusieurs traductions et interprétations, ils arrivent souvent sous une forme dégradée et difficilement compréhensible.

Même si la traduction automatique s'améliore rapidement, il est peu probable que des machines puissent jamais atteindre le degré de nuance et d'exactitude technique que requièrent les transformations constantes auxquelles est soumis le vocabulaire scientifique.

La mondialisation de la science offre sans conteste de très nombreuses opportunités en matière d'avancées intellectuelles. Mais à moins que des passerelles plus solides ne relient les différentes communautés

linguistiques entre elles, il est à craindre qu'une quantité incalculable d'idées et d'innovations ne soient ignorées et finissent par se perdre.

Ce qui est vrai de la mondialisation de la communication scientifique [LÉVY-LEBLOND 2007] l'est, à plus forte raison, des autres domaines : le cyberspace ne saurait se contenter d'une seule *lingua franca*, pour puissante que soit encore la langue anglaise. Il suffit de naviguer sur internet pour se rendre compte de l'avantage immense que représente le regard du plurilingue : ne voir le monde que par l'intermédiaire de l'anglais offre, par contraste, une vision de plus en plus lacunaire au fur et à mesure que les autres langues progressent aussi bien en termes absolus que relatifs. C'est là un changement de paradigme dont on commence seulement aujourd'hui à mesurer toute l'ampleur. La communication n'est plus pensable à l'échelle d'une langue et d'une seule : il faut lui adjoindre le plurilinguisme et la traduction (y compris automatique ou assistée par ordinateur). Ce sont là, néanmoins, des conditions nécessaires, mais non suffisantes.

« *La guerre est une affaire trop sérieuse pour être confiée à des militaires* », disait Clémenceau. Il en va de même de la question des langues que ce soit sur le cyberspace ou dans le cadre plus vaste d'un monde en voie de reabélisation accélérée : elle ne saurait être la chasse gardée des linguistes, des traducteurs et des interprètes ou des traductologues. Elle est du ressort d'une approche véritablement pluridisciplinaire faisant appel aussi bien aux sciences humaines et aux humanités qu'aux sciences de la nature ou dites « dures » [OUSTINOFF 2011]. Plus généralement, elle est désormais l'affaire de tous puisque le cyberspace ne s'adresse pas, loin s'en faut, qu'aux seuls spécialistes.

BIBLIOGRAPHIE

[BARANY 2005] Barany, M. J., *Science's Language Problem*, Business Week, 16 mars 2005. http://www.businessweek.com/technology/content/mar2005/tc20050317_4179.htm

[BRITISH ACADEMY 2009] *The British Academy, Language Matters. Position Paper*, 2009. <http://www.britac.ac.uk>

[BEL HABIB 2011] Bel Habib, I., *Multilingual Skills Provide Export Benefits and Better Access to New Emerging Markets*, Sens Public, octobre 2011. http://www.sens-public.org/article.php3?id_article=869

[CALVET 2007] Calvet, L.-J., *La traduction au filtre de la mondialisation*, dans Oustinoff, M., Nowicki, J. (dir.), *Traduction et mondialisation*, Hermès, n°49, Paris, CNRS Éditions, 2007.

[CASSIN 2004] Cassin, B. (dir.), *Vocabulaire européen des philosophies. Dictionnaire des intraduisibles*, Paris, Le Robert / Le Seuil, 2004.

[CRYSTAL 1997] Crystal, D., *English as a Global Language*, Cambridge, Cambridge University Press, 1997.

[GRADDOL 2000] Graddol, D., *The Future of English ? A Guide to Forecasting the Popularity of the English Language in the 21st Century*, The British Council & The British Company (UK) Ltd, 1997 (nlle éd., 2000). <http://www.britishcouncil.org/learning-elt-future.pdf>

[MARTEL 2010] Martel, F., *Mainstream. Enquête sur cette culture qui plaît à tout le monde*, Paris, Flammarion, 2010.

Multilingual WorldWideScience.org. <http://worldwidescience.org/multi/index.html>

[LÉVY-LEBLOND 2007] Lévy-Leblond, J.-M., *Sciences « dures » et traduction*, dans Oustinoff, M., Nowicki, J. (dir.), *Traduction et mondialisation*, Hermès, n°49, Paris, CNRS Éditions, 2007.

[OUSTINOFF, NOWICKI 2007] Oustinoff, M., Nowicki, J. (dir.), *Traduction et mondialisation*, Hermès, n°49, Paris, CNRS Éditions, 2007.

[OUSTINOFF, NOWICKI, MACHADO DA SILVA 2010] Oustinoff, M., Nowicki, J., Machado da Silva, J. (dir.), *Traduction et mondialisation. Volume 2*, Hermès, n°56, Paris, CNRS Éditions, 2010.

[OUSTINOFF 2011] Oustinoff, M., *Traduire et communiquer à l'heure de la mondialisation*, Paris, CNRS Editions, 2011.

[PAOLILLO, PIMIENTA, PRADO 2005] Paolillo, J., Pimienta, D., Prado, D., et al., *Mesurer la diversité linguistique sur Internet*, révisé et accompagné d'une introduction de l'Institut de statistique de l'Unesco, Publications de l'Unesco pour le Sommet mondial sur la société de l'information, Paris, 2005. <http://unesdoc.unesco.org/images/0014/001421/142186f.pdf>

[QUIRK, AL. 1980] Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., *A Grammar of Contemporary English*, Londres, Longman, 1980.

[RICARD 2007] Ricard, P *Une étude britannique prône le multilinguisme en affaires*, Le Monde, 25 septembre 2007.

[SHIN, KOMINSKI 2010] Shin, H. B., Kominski, R. A., *Language Use in the United States : 2007. American Community Survey Reports*, Washington D.C., US Department of Commerce. Economics and Statistics Administration, US Census Bureau, 2010. <http://www.census.gov/prod/2010pubs/acs-12.pdf>

[WOLTON 2003] Wolton, D., *L'autre mondialisation*, Paris, Flammarion, 2003.

INNOVATION TECHNOLOGIQUE ET MAINTIEN DES LANGUES

Quand l'industrie des nouvelles technologies innove, elle est principalement guidée par une vision d'affaire à court et moyen terme. Les langues au contraire s'évaluent sur le long terme. Comment fonctionne cette rencontre d'échelles de temps ? Ne faudrait-il pas penser l'innovation des TIC autour de la diversité linguistique ?

Article original en français



ÉRIC PONCET a fondé Linguasoft pour accompagner les communautés désireuses de maintenir leur langue. Concepteur d'une méthode et d'outils spécifiques de préservation des langues, il a encadré de nombreux projets en ce sens.

ÉRIC PONCET

DES LANGUES
ET MAINTIEN
TECHNOLOGIQUE
INNOVATION

Cet encart présente l'influence des technologies de l'information et de la communication (TIC) sur l'évolution des langues du monde, et argumente en faveur d'une innovation de ces TIC centrée sur le multilinguisme.

Langues, technologie, et temps

L'écriture peut être considérée comme la plus ancienne des technologies langagières (ne serait-ce que parce qu'aucune autre avant elle n'a laissé de trace). Pourtant, ses cinq millénaires d'existence ne sont qu'une infime période de la vie du langage humain. Pour mémoire, le crâne d'*Homo habilis* (2,5 millions d'années) contenait l'aire de Broca, zone du langage. Si cela ne signifie pas que le langage humain a 2,5 millions d'années, cela donne au moins une idée de la jeunesse de la technologie au regard du langage. Or l'écriture, malgré ce jeune âge, a eu une influence aussi majeure que rapide sur l'évolution des sociétés humaines. C'est précisément ce phénomène de révolution technologique éclair – à l'échelle de notre espèce – que nous vivons actuellement avec l'internet.

Les chercheurs et activistes en langues s'accordent à dire que la moitié des 6900 langues actuelles disparaîtront d'ici un siècle. Il est probable que cette opinion sous-estime – et dans d'importantes proportions – l'impact présent et à venir du cyberspace sur le multilinguisme.

Linguicide technologique

L'innovation en nouvelles technologies étant – du moins dans l'industrie – souvent guidée par une vision d'affaires à court et moyen terme, les TIC sont au regard des langues du monde une arme à double tranchant. On

conçoit aisément qu'un éditeur de logiciels localisant son produit grand public pour le marché chinois vise à pénétrer un marché de quasiment un milliard de locuteurs en mandarin, sur lequel il pourra distribuer des millions de licences. Que se passe-t-il si cette même société veut localiser ce même produit pour une langue comptant moins de 10 000 locuteurs (soit la moitié des langues du monde actuellement) ? D'une part, le nombre de licences sera de quelques dizaines ou centaines, et d'autre part, son travail de localisation sera plus long (donc plus coûteux) que pour le mandarin. En effet, nombre de ces langues ne sont pas normalisées, voire n'ont pas de système d'écriture. Au final, l'éditeur devra investir significativement plus et accepter de gagner beaucoup moins. Quelle entreprise est aujourd'hui capable (ou simplement volontaire) pour convaincre ses actionnaires et employés qu'une telle démarche est à considérer dans sa stratégie d'affaires ? Et que dire enfin de la faisabilité d'une localisation d'un logiciel en 6900 langues ? Il est à noter qu'aucun des millions de logiciels existant de par le monde ne se rapproche – même de loin – de ce niveau de multilinguisme.

Innovation et multilinguisme

Alors... avec les technologies, point de salut ?

Le multilinguisme actuel des technologies n'est que trop restreint, notamment parce qu'il se limite, pour les raisons évoquées plus haut, aux langues les plus puissantes. Il convient d'étendre ce multilinguisme vers un panlinguisme. Autrement dit, plutôt que de *saupoudrer* les technologies de langues, il devient nécessaire de les repenser pour y *intégrer* le langage – en tant que caractéristique forte de l'humanité.

Malheur aux langues qui n'ont pas la masse critique, tant en nombre de locuteurs qu'en financement. Leur faible présence sur la Toile – si ce n'est leur absence – a pour conséquence que leurs locuteurs sont, par défaut, linguistiquement absorbés par les langues présentes. L'analogie astrophysique avec un trou noir n'est pas trop forte : les langues n'ayant pas l'inertie ou la masse suffisante pour poursuivre leur trajectoire se feront irrémédiablement absorber, et ce d'autant plus vite que leur masse est faible.

Si le premier aspect de la masse critique (nombre de locuteurs) ne peut être augmenté du jour au lendemain par un simple coup de baguette

magique, le deuxième (financement) peut en revanche l'être : des projets de maintien des langues peuvent être lancés avec de modestes budgets, et tout financement en ce sens est un catalyseur d'énergies fécondes dans les communautés linguistiques intéressées.

Prospective

Revenons sur le taux de disparition des langues d'ici un siècle mentionné plus haut. L'auteur, au vu des tendances actuelles, et du travail de terrain qu'il a effectué, penche plutôt pour un taux de disparition entre 80 et 95%.

S'il a fallu à l'écriture plusieurs siècles pour conquérir le monde, il n'a fallu qu'une décennie à internet pour révolutionner la façon dont ses utilisateurs communiquent, consomment, travaillent, se divertissent... bref, la façon dont ils vivent. Il est important de noter que c'est bien le langage (et les langues) qui est le vecteur de toutes ces activités. Le réseau des réseaux étant un vecteur linguistique majeur, que peut-on présager pour le siècle à venir – soit dix fois l'âge actuel de l'internet ? Et peut-on exclure l'apparition d'une technologie dont la diffusion serait plus foudroyante encore que l'écriture ou l'internet ? Un tel événement ne laisserait à la majorité des langues du monde que peu de chances de survie, et pourrait engendrer à (court) terme un cataclysme culturel.

« *Science sans conscience n'est que ruine de l'âme* » écrit François Rabelais. Cette citation visionnaire, après avoir brillamment traversé les cinq derniers siècles, n'est que trop facile à transposer « musicalement » au thème du présent ouvrage : TIC sans éthique n'est que ruine de l'Homme.

PRÉSERVER L'HÉRITAGE DES LANGUES ÉTEINTES OU EN DANGER

Cet article a pour ambition de montrer comment le cyberspace peut soutenir la préservation des langues éteintes ou en danger d'extinction. Différentes études de cas montrent comment le cyberspace a pu remplir ce rôle et pour quelles raisons les langues concernées sont ainsi préservées. Nous ne privilégions pas une approche globalisante de la question, puisque des facteurs aussi différents que l'objectif de la préservation, les attitudes à l'égard de la langue au sein des communautés, l'existence d'une diaspora ou encore l'accessibilité aux technologies numériques auront chacun un rôle à jouer.

Article original en anglais.
Traduit par Laurent Vannini.



MAIK GIBSON est directeur du département des Études en Traduction à l'Université internationale de Nairobi, Kenya, et consultant international de sociolinguistique au sein de SIL International. Il est également professeur de sociolinguistique à i-Delta, Institut pour le Développement des Langues et de la Traduction en Afrique. Ses recherches portent entre autres sur la conversion linguistique, l'aménagement linguistique et l'expansion des domaines de pratique des langues moins répandues.

MAIK GIBSON

PRÉSERVER
L'HÉRITAGE
DES LANGUES
ÉTEINTES OU
EN DANGER

Le vingt et unième siècle est témoin de l'émergence de deux tendances, chacune reflétant le monde en mouvement dans lequel nous vivons. La première est la propagation de l'internet et des technologies de l'information et de la communication ; la seconde est un déclin rapide du nombre de langues parlées et de celles qui peuvent prétendre à un avenir. Nous nous intéresserons dans le présent chapitre à la manière dont internet peut être utilisé comme outil efficace de prévention et d'atténuation des risques de disparition et d'extinction des langues.

Il existe de nombreuses estimations relatives au nombre de langues menacées d'extinction, dont les plus pessimistes évaluent un risque pour 50 % d'entre elles. À titre d'exemple, l'*Atlas Unesco des langues en danger dans le monde* [MOSELEY 2010] donne le chiffre « d'environ 3 000 langues » concernées. Mais quel que soit le chiffre exact, il est évident que l'extinction des langues sera une tendance majeure du XXI^e siècle, et ce principalement dans les Amériques et en Australie. Lorsqu'une langue meurt, les savoirs des régions associées peuvent disparaître à leur tour et la communauté touchée risque de perdre l'un des symboles essentiels de son identité ethnique, ce qui pourrait être facteur d'instabilité sociale. Il semble évident qu'un grand nombre de langues maternelles ne seront pas la première langue d'usage de la prochaine génération d'enfants. Cependant, la récolte et l'enregistrement de matériaux illustrant l'actuelle diversité peuvent aider la communauté à maintenir ses savoirs traditionnels, et l'usage de sa langue, ne serait-ce qu'à titre patrimonial, et contribuer ainsi à résoudre des conflits ou questions identitaires. La présence d'internet rend la tâche de collecte de ressources moins ardue, et permet également d'accéder plus facilement à cette documentation – non seulement pour la communauté concernée, si toutefois elle dispose d'une connexion, mais aussi pour toute personne ou groupe intéressés. Le but de cet article est

de fournir des clés de compréhension et de décrire les différents facteurs pouvant rendre l'internet utile à la récolte et l'enregistrement de données sur des langues menacées d'extinction ou éteintes.

LANGUES ÉTEINTES ET LANGUES EN VOIE DE DISPARITION

Les langues éteintes (autrement appelées langues *mortes*) et en voie de disparition sont situées au plus bas de l'échelle de vitalité linguistique, qui peut être évalué à l'aide d'outils comme la méthodologie pour évaluer la vitalité et le danger de disparition des langues développée par l'Unesco [BRENZINGER 2003], l'échelle graduée de discontinuité intergénérationnelle (GIDS) de Fishman [FISHMAN 1991] ou encore sa version étendue proposée par Lewis et Simons en 2010 [LEWIS, SIMONS 2010]. Les mesures de la vitalité linguistique sont en premier lieu sensibles au présent et au futur probable d'une langue comme véhicule de la socialisation primaire des enfants. Dans un cas extrême, celui de l'extinction, la situation est celle d'une langue qui n'est plus utilisée par quiconque dans les relations au quotidien. Dès lors, les enfants n'ont plus l'opportunité de l'apprendre comme langue première que dans le contexte d'une revitalisation délibérée. Extinction rime souvent avec perte de toute connaissance de la langue, mais ce n'est pas toujours le cas – les enfants et grands-enfants des derniers locuteurs ayant la maîtrise d'une langue en ont parfois une connaissance passive du fait de l'avoir entendu parler dans leurs premières années, ou peuvent même s'exprimer au travers de quelques mots ou phrases.

Dans d'autres cas, si la langue a pu disparaître des conversations courantes, son usage peut subsister lors de pratiques ou rituels religieux. C'est par exemple le cas du ge'ez, du latin ou encore du vieux-slave liturgique, qui sont des exemples de langues éteintes qui ont été écrites. Mais des scénarios similaires existent pour des langues sans système d'écriture, comme le lucumi à Cuba, que décrit Courlander dans *A Treasury of AfroAmerican Folklore* [COURLANDER 1996: 20]. Aussi, le terme *éteinte* (tout comme son synonyme contextuel de *morte*) ne signifie pas forcément qu'une langue est définitivement perdue, ou qu'elle n'a plus aucune fonctionnalité dans la communauté, mais indique le manque flagrant de locuteurs pour lesquels elle serait l'un des premiers moyens de communication. De toutes évidences, une langue éteinte dont la moindre parcelle de savoir

a été perdue, ne pourra pas être préservée. Un niveau minimum de connaissance, imprégnant la communauté ou sous la forme de ressources documentaires, est un pré-requis.

Si l'évaluation du nombre restreint de locuteurs premiers, ou *maternels*, d'une langue suffit, en théorie, à la désigner comme éteinte, il est plus compliqué de déterminer si elle est en voie de disparition. En général, une langue qui n'est pas transmise aux enfants sera automatiquement considérée comme en voie de disparition, puisque l'extinction est l'aboutissement le plus fréquent après deux générations. Il existe toutefois des cas exceptionnels où la langue de la communauté n'est pas le premier idiome enseigné aux enfants, mais sera acquise pleinement à un âge plus avancé. Ainsi, pour le digo au Kenya, Nicolle [NICOLLE 2012 : 4,5] indique : « *la majorité des jeunes enfants entendaient leurs parents et d'autres adultes leur parler en swahili... pourtant, à l'âge de 10 ou 11 ans, ils étaient tous quasiment devenus des locuteurs expérimentés du digo et les adultes leur adressaient usuellement la parole en digo.* » En outre, des situations dans lesquelles seule une fraction des enfants apprendrait la langue de la communauté peuvent dégénérer en disparition – c'est ce qui se passe habituellement lorsqu'il y a déjà un risque d'extinction. Ce risque peut être difficile à quantifier dans chaque cas, la conversion linguistique pouvant être ralentie, stoppée ou même inversée. Par ailleurs, Moseley [MOSELEY 2010] classe les langues qui ne sont pas utilisées en toutes circonstances (le gallois par exemple) comme vulnérables. Un usage circonstanciel qui est pourtant la norme dans les sociétés multilingues, où les différentes langues sont assignées à des fonctions distinctes, qu'une conversion linguistique soit en cours ou non.

LA DIVERSITÉ LINGUISTIQUE S'ÉTIOLE PARTOUT DANS LE MONDE

L'extinction des langues n'est pas exclusivement un phénomène moderne, même si nous pouvons affirmer qu'il s'est fortement accéléré en raison des changements sociaux abrupts accompagnant la globalisation depuis un siècle. Même des langues à forte tradition écrite comme le sumérien, le vieux-prussien ou encore le massachusett se sont éteintes sans que subsiste aujourd'hui la moindre descendance linguistique. Dans le même temps, un nombre incalculable de langues, dont il n'a existé que peu ou

pas de trace écrite, ont disparu, ne laissant épisodiquement dans notre présent que d'obscurs noms de lieux. Il nous est quasiment impossible, par manque de documentation, de comparer l'effritement de la diversité linguistique au xx^e siècle avec celui des siècles précédents.

Malgré tout, il semble que la prédiction d'un scénario extrême, soit une disparition dans les cinquante à cent années à venir de 50 % des 6 000 langues composant le monde linguistique, représente la vague de réduction de la diversité la plus importante de tous les temps. Les conséquences sont de plusieurs types. Souvent, la conversion linguistique s'accompagne d'un changement de style de vie, provoquant ainsi la perte simultanée de savoirs et de vocabulaires traditionnels, dans le champ de l'ethnobotanique par exemple, où les langues sont porteuses d'une distinction entre plantes que la science moderne n'a pas étudiées. Lorsque les changements de style de vie sont minimes, le vocabulaire traditionnel survit parfois. Brenzinger [BRENZINGER 1992] en donne deux exemples au Kenya – le yaaku et l'elmolo, dont les locuteurs ont remplacé progressivement leur dialecte communautaire par des dialectes dérivés du maa. Les deux langues ont cependant préservé des aspects de leurs différents styles de vie pour lesquels le maa ne contenait que peu de vocabulaire (l'apiculture et la chasse pour le yaaku, la pêche pour l'elmolo), et ont conservé les termes relatifs à ces domaines. Aujourd'hui, les anciens yaaku adoptent le style de vie de la culture pastorale Maasai, perdant ainsi leur vocabulaire propre à la chasse et à l'apiculture. Même dans des cas où les styles de vie sont maintenus, une compréhension du sens et de l'origine de termes spécialisés peut disparaître lorsque la langue dont ils sont issus est oubliée. Harrison, dans son ouvrage *When Languages Die* [HARRISON 2007] donne un aperçu plus complet des types de connaissances et de visions du monde qui peuvent disparaître.

Autre phénomène découlant de l'extinction d'une langue indigène, on assiste parfois à une rupture culturelle, qui se traduit par des taux de suicide beaucoup plus importants dans les communautés dont la langue a été perdue ou est en passe de l'être [HALLETT 2007]. En Colombie Britannique, la perte d'une langue communautaire est la corrélation la plus significative aux taux de suicide élevés constatés dans les communautés autochtones, tous les autres « facteurs de continuité culturelle » étudiés présentant des degrés de corrélation moindres. Si un tel schéma de rupture culturelle ne se reproduit pas dans tous les cas de disparition de langue communautaire,

cette étude montre que la perte d'une langue ne s'opère pas au milieu d'un néant social, et peut avoir des impacts dramatiques sur la communauté concernée.

La grande diversité des langues qui sont encore parlées aujourd'hui illustre une incroyable variété de stratégies mises en œuvre pour exprimer la pensée humaine et catégoriser le monde autour de nous. Aussi, d'un point de vue scientifique, l'extinction d'une langue signifie une diminution des données disponibles pour comprendre les fonctionnements de l'esprit humain, non seulement au regard de faits et théories linguistiques, mais également à propos des capacités générales de l'esprit humain, cognitives ou perceptives par exemple. Et chaque langue prise individuellement est d'importance équivalente pour développer notre savoir dans ces domaines. Fonder une théorie universelle de la langue ou de la cognition sur une poignée de langues seulement n'est pas pertinent. Chaque langue éteinte représente une opportunité perdue pour améliorer la connaissance que nous avons de l'être humain.

EXEMPLES D'INTERVENTIONS PRENANT APPUI SUR L'INTERNET

Le choix d'un type de préservation, dans le cyberspace ou non, dépendra de ses objectifs. Si le but est simplement de préserver la langue pour la postérité, quelques textes, du vocabulaire et des enregistrements publiés sur internet feront sans doute l'affaire. S'il est décidé que la langue puisse être apprise comme seconde langue par la génération actuelle et celles futures, il sera alors nécessaire de disposer d'une base suffisante pour développer du matériel d'enseignement adapté, et une classe interactive en ligne pourrait être instaurée pour celles et ceux souhaitant apprendre la langue. Si la langue est toujours au cœur de certaines relations de la communauté, comme pour l'accomplissement de rituels, le développement d'outils et documents correspondants pourrait être une solution, dans le cas où la société concernée ne considère pas la langue de rituel comme quelque chose de secret, dont la connaissance ne doit pas être partagée en dehors d'un groupe précis de personnes.

Des actions de préservation visant à revitaliser une langue requerront des efforts plus importants. Cependant quelques interventions simples, en ligne, comme la mise en place de sites internet et de forum sur lesquels

les locuteurs sont encouragés à utiliser la langue menacée, peuvent être très efficaces selon le profil social des personnes impliquées.

L'utilité de l'internet comme outil pour protéger les langues en voie de disparition est intelligemment démontré par Mensching [MENSCHING 2000], dans une étude sur le sarde, une langue romane caractérisée par de fortes différences entre les différents dialectes. En l'occurrence, Moseley [MOSELEY 2010] distingue et sépare quatre branches qu'il considère toutes comme en voie de disparition, tandis que de son côté Lewis, dans la seizième édition de *Ethnologue: Languages of the World* [LEWIS 2009] liste ces quatre branches comme appartenant à la macro-langue sarde, cumulant dès lors un nombre de locuteurs de plus d'un million de personnes. D'une certaine manière, nous avons ici un terrain idéal pour la réussite du projet, au regard du nombre important de locuteurs utilisant la langue aujourd'hui et d'un contexte d'une société fortement développée avec des revenus relativement élevés, un accès internet et un bon niveau d'éducation. Le principal indicateur négatif étant le haut niveau de disparité entre les différents dialectes, mais Mensching avance l'argument que le médium de l'écriture minimise ces différences, puisque dès lors les locuteurs convergent vers un système commun d'écriture du sarde. Il mentionne plusieurs atouts propres à l'internet, en particulier la présence d'un site internet dédié à la culture et à la langue sarde¹, dans le sens où cela :

- « *renforce la conscience linguistique des locuteurs* », en partie parce que le sarde n'est pas seulement le sujet du site dédié, mais également son médium, donnant ainsi aux utilisateurs une pratique de la forme écrite ;
- aide les locuteurs en leur donnant de l'information sur la manière d'écrire la langue, rendant ainsi la forme écrite plus homogène, et donc plus accessible pour toute la communauté ;
- agit comme « *un nœud central pour accéder à de l'information en ligne sur la langue sarde et la culture de l'île* » ;
- sert d'espace de discussion, de documentation et d'évaluation, tout cela se déroulant sans la médiation d'un langage dominant (situation qu'il est commun de rencontrer dans les activités académiques) ;

1 <http://www.lingrom.fu-berlin.de/SardischEngl.html>

- prend avantage de « l'effet d'oralité » spécifique à l'internet, dans lequel « *la communication par e-mails est plus proche de la langue orale que de la langue écrite* ». Cette observation peut être effectuée pour d'autres langues minoritaires et des dialectes « non chapeautés² », qui ont une présence significative dans les écrits en ligne, d'autant plus aujourd'hui avec le développement du Web 2.0 et d'applications comme les sites de mise en réseau, à l'instar de Facebook³. La catégorisation habituelle de l'écriture comme un domaine formel, où la langue dominante prédomine, semble ne pas s'appliquer à certaines formes d'écrits en ligne, qui ne sont pas sans rappeler la langue déployée pour rédiger des SMS à partir de téléphones mobiles.

Ces facteurs jouent sans doute un rôle dans le succès rencontré sur l'internet par l'activisme en lien aux langues minoritaires. Nous pouvons dès lors formuler l'hypothèse que des bénéfices similaires sont à l'œuvre pour des langues minoritaires situées dans des contextes similaires (en Europe de l'Ouest par exemple), où l'on retrouve les mêmes critères d'alphabétisation élevée et d'accès à l'internet, ainsi qu'une ressemblance entre les systèmes d'écriture et grammaires des langues dominantes et minoritaires. Les sites internet permettent également aux membres de la diaspora, qui ne vivent pas dans la région où se parle leur langue, de la pratiquer, là où auparavant cette pratique aurait été restreinte aux seules conversations téléphoniques. De même, ces migrants ne perdent pas leurs compétences linguistiques et peuvent contribuer activement à la préservation de leur langue⁴. Néanmoins, tous ces facteurs ne seront pas aussi féconds dans des contextes où la langue en voie de disparition est parlée par une minorité marginalisée, dans des régions caractérisées par de plus faibles taux d'alphabétisation et un accès restreint aux nouvelles technologies, ou encore là où existent des différences significatives au cœur des systèmes d'écriture et de la structure des langues, particulièrement dans l'organisation des sons – la phonologie.

Là où le contexte diffère, les stratégies de revitalisation d'une langue via la Toile peuvent également diverger. Dans le cas de certaines communautés

2 Expression utilisée par Heinz Kloss et Grant D. McConnell dans leur livre *Linguistic Composition of the Nations of the World: Europe and the USSR* pour désigner des dialectes non immédiatement reliés à une langue standard.

3 Voir dans cet ouvrage : Vassili Rivron, *L'usage de Facebook chez les Éton du Cameroun*.

4 Voir dans cet ouvrage : Viola Krebs & Vicent Climent-Ferrando, *Langues, cyberspace et migrations*.

Natives Amérindiennes, alors qu'une conversion linguistique importante est en cours, les sites internet sont utilisés dans le cadre d'une stratégie plus large de transmission de la langue aux nouvelles générations (incluant entre autres des camps d'été linguistiques, durant lesquels les enfants sont encouragés à utiliser la langue de la communauté entre eux). Des sites de ce type existent par exemple pour la langue anishinaabemowin⁵ ou pour des cours en ligne en potawatomi⁶. Étant donné que l'anglais est la langue première pour la majorité des membres de ces communautés, les sites internet sont proposés en anglais à la connexion, et servent de passerelles vers les langues qu'ils soutiennent. Un choix qui contraste avec celui du site internet dédié à la langue sarde, et qui n'utilise pas la langue dominante. Cette stratégie en ligne peut avoir un impact considérable, en raison de la forte pénétration d'internet au sein de ces communautés amérindiennes. Galla [GALLA 2009] propose une analyse pertinente des problématiques liées aux technologies de pointe dans des contextes similaires à celui de l'Amérique du Nord.

La planète est de plus en plus connectée à Internet, bien qu'il existe toujours des problèmes saillants de vitesse et de facilité d'accès notamment dans les pays en voie de développement, et plus particulièrement encore dans les contextes ruraux où se parlent généralement les langues en voie de disparition. La pénétration progressive des téléphones mobiles dans les campagnes kényanes par exemple, est un signe encourageant, et qui va de pair avec la réduction du coût des téléphones permettant de naviguer sur le Web. Pour cette raison, l'accès à internet n'est plus dépendant d'une alimentation permanente en électricité ou de câbles pour le haut-débit, sitôt que des solutions sont trouvées pour recharger le téléphone, en utilisant par exemple une batterie de voiture, un vélo ou un système utilisant l'énergie solaire. Cette nouvelle technologie peut donc soutenir les efforts fournis pour revitaliser une langue, comme l'a démontré K. David Harrison dans son dictionnaire en ligne du touvain⁷, une langue turque.

En Afrique, la majorité des langues en voie de disparition sont parlées par des chasseurs-cueilleurs, petite minorité au sein de la population globale, et qui dans la plupart des cas sont peu respectés par les communautés voisines. Par exemple, au Kenya, selon les travaux de Brenzinger [BREZINGER

5 <http://anishinaabemdaa.com>

6 <http://www.potawatomilanguage.org/revitalization.php>

7 <http://www.swarthmore.edu/SocSci/tuvan/dict/> Celui-ci est également téléchargeable gratuitement sous la forme d'une application pour l'iPhone depuis les boutiques iTunes.

1992: 215] «*Les chasseurs sont considérés par les éleveurs de bétail comme 'pauvres', 'primitifs', 'vivant comme des animaux', etc.* ». Aussi, lorsque des groupes de chasseurs-cueilleurs entrent en relation symbiotique avec les populations voisines, on observe couramment une tendance à l'abandon de leur langue et de leur style de vie [DIMMENDAAL 1989]. Toutefois, ce style de vie, quelle que soit la langue, se caractérise par des niveaux très bas d'alphabétisation et de contact avec l'internet. Dès lors, celui-ci s'avère peu efficace comme point d'appui pour la préservation de la langue dans un rapport direct à la communauté – d'une certaine manière, l'extrême opposé du scénario de la langue sarde. Cependant, si les conditions culturelles le permettent, poster sur internet des ressources audio et vidéo propres à ces langues est toujours utile à la fois pour les membres des communautés concernées mais également pour des linguistes ou d'autres personnes intéressées. Une démarche de préservation des aspects de la langue pour la postérité, qui ne préjuge pas de l'usage qu'en fera la communauté concernée dans le futur⁸.

Nous avons présenté trois types de scénarios différents :

- celui du sarde, parlé par une population importante de locuteurs dotés d'un bon niveau d'éducation, d'un large accès à internet et n'ayant pas de difficultés majeures à y employer leurs compétences écrites de l'italien. Ici, l'usage du sarde dans une gamme d'activités ancrées sur l'internet ne pose que peu de problèmes ;
- les contextes d'un grand nombre de langues amérindiennes d'Amérique du Nord, qui partagent les mêmes conditions fécondes, à savoir des niveaux d'éducation élevés et un accès à internet. Pour ces langues, il ne subsiste souvent qu'un groupe restreint de locuteurs, la majorité de la population ayant adopté l'anglais comme langue première. Il s'agit cependant d'une population dont les membres peuvent avoir un jour envie d'apprendre leur langue communautaire. Il est à souligner également que l'écart linguistique entre les langues des Natifs Amérindiens et l'anglais est plus large qu'entre l'italien le sarde. Aussi, la priorité dans ces contextes est à donner au soutien à l'apprentissage de la langue, en parallèle à une mise en avant primordiale de la culture communautaire ;

⁸ Voir dans cet ouvrage : Tjeerd De Graaf, *Intérêt des archives orales pour les langues menacées*.

- dans le contexte des groupes de chasseurs-cueilleurs d’Afrique, aucun des facteurs susmentionnés, favorisant un activisme linguistique sur internet, n’existe. Dans ce cas, des initiatives devront sans doute être prises par des tierces personnes militant pour la préservation de ces langues, via l’enregistrement d’échantillons de leur utilisation par exemple, dans la mesure où ces initiatives sont jugées favorablement par la communauté elle-même. Dans de telles conditions, des groupes de discussion ou du matériel pédagogique en ligne n’auront que peu de valeur, tant que les questions d’accès à internet et de motivation communautaire à préserver la langue n’auront pas été préalablement résolues. Dans certains cas, aucune forme écrite de la langue n’existe, et bien que cela ne constitue pas en soi un indicateur du risque de disparition encouru par l’idiome, le développement d’un système d’écriture en plus de la constitution de matériaux écrits, peut aider à l’expansion des domaines parcourus par la langue, augmentant son prestige de telle manière que se réduisent les motivations sociales pour basculer vers une autre langue⁹. Ces trois scénarios ne sont en aucun cas exhaustifs, mais démontrent que les interventions prenant appui sur internet pour la préservation de langues doivent être à l’écoute de considérations sociales et linguistiques si elles veulent avoir une chance d’inverser la tendance contemporaine à la disparition des langues.

CHOISIR LE TYPE DE DOCUMENTATION APPROPRIÉ

Au regard de la menace d’une extinction massive de langues pesant aujourd’hui sur nos épaules, préserver ce qui sera ou pourrait être perdu est une priorité sans commune mesure pour les linguistes et les personnes intéressées par la sauvegarde des cultures afférentes. S’il paraît évident que les linguistes professionnels ont un rôle à jouer, ils ne sont pas les seuls à pouvoir influencer positivement le cours de l’histoire. La préservation peut prendre diverses formes, parmi lesquelles on trouve évidemment la grammaire et le dictionnaire chers au linguiste, mais aussi des gestes de nature moins académiques comme l’enregistrement de séquences audio et vidéo témoignant de l’usage de la langue, ou encore d’un recueil d’histoires ou de récits issus des savoirs locaux, qui sont au moins aussi

⁹ Voir dans cet ouvrage : Evgeny Kuzmin, *Politiques linguistiques pour contrer la marginalisation des langues*.

importants. De surcroît, la mise en place d'un dictionnaire complet peut s'avérer un énorme challenge logistique dans certaines situations, alors que des projets aux ambitions plus restreintes (comme l'établissement d'un lexique ou d'un glossaire couvrant des aspects particuliers d'une langue, ou une brève introduction à sa grammaire) peuvent être plus faciles à finaliser, et ont l'avantage supplémentaire d'être accessibles à une plus grande audience. Des outils internet existent, tels que WeSay¹⁰, pour aider des « *non-linguistes à établir un dictionnaire dans leur propre langue* ». S'engager dans de telles démarches sur l'internet réduit également les coûts et difficultés logistiques inhérents à la production et à la distribution de livres. Néanmoins, dans un grand nombre de cas pour lesquels l'accès aux technologies est limité, le livre imprimé reste parfois la meilleure solution. Toute intervention doit donc prendre en considération l'utilisation de technologies appropriées, ainsi que les valeurs sociales rattachées à la langue, à l'alphabétisation et à l'appropriation de matériels.

Ainsi, pour prendre l'exemple du munichi, une langue du Pérou désormais éteinte, et qui ne compte plus qu'une demi-douzaine de « sous-usagers »¹¹, la documentation linguistique existante [GIBSON 1996] a été publiée sur le Web par SIL International. Il est ainsi disponible gratuitement aux membres de la communauté. Cependant, de telles études, destinées à un usage par des linguistes plus que par des membres de la communauté elle-même, peuvent être difficiles d'accès pour celles et ceux qui n'ont pas une formation en linguistique ; le vocabulaire et les exemples seront appréciés et utiles, mais dans l'ensemble, la terminologie linguistique vise à la précision, et n'est pas vraiment efficace pour les néophyte. Fort heureusement pour le peuple des munichis, un projet de recherche additionnel a été initié [MICHAEL 2009] avec comme objectif de rendre disponible des ressources pour la communauté, notamment des échantillons audio. De tels matériels destinés à la communauté, qu'ils soient en ligne ou disponibles sur d'autres médias, devraient idéalement prendre une forme différente de ceux adressés en premier lieu à la communauté scientifique – le passage d'une forme à une autre étant possible dans la plupart des cas. Les matériaux premiers (enregistrement ou transcriptions de récits, rituels, conversations etc.) sont utiles à la fois aux membres de la communauté et également aux personnes qui veulent approfondir des travaux d'analyse et de développement de la langue, et ne doivent par conséquent pas être

¹⁰ <http://www.wesay.org>

¹¹ Claude Hagège, *Halte à la mort des langues* éd. Odile Jacob, Paris, 2000.

remplacés par des traitements grammaticaux. Après tout, les enregistrements d'une langue parlée sont les sources premières du linguiste.

En conclusion, si la tentation existe d'utiliser l'anglais comme *lingua franca* de l'internet, ou même de voir internet comme un moteur d'uniformisation linguistique plutôt qu'un facteur de diversité, on doit constater néanmoins que le cyberspace peut faciliter, et facilite déjà, la documentation et l'utilisation de langues en voie de disparition, qui peuvent alors s'emparer d'un espace considéré comme une facette de la mondialisation pour y semer les graines de survie de la diversité linguistique. L'internet est parfaitement adapté au dépôt de matériaux linguistiques, quelles que soient leurs formes. Et ceci indépendamment des objectifs de chaque projet, entre préservation, incitation à l'utilisation, pédagogie etc. Comme nous avons pu le voir, la manière dont internet peut contribuer à la préservation de la diversité peut revêtir de nombreuses formes, dépendantes pour cela de facteurs sociaux, technologiques ou linguistiques, qui eux-mêmes sont dynamiques, soumis aux évolutions des technologies et des attitudes communautaires. Par conséquent, nous ne pouvons plaider pour une approche de type « unique et unilatérale » d'utilisation d'internet comme ressource dans la lutte contre l'extinction ou le risque de disparition des langues. Mais nous encourageons l'adoption de formes d'interventions qui correspondent au mieux à l'écologie de la langue ciblée, dans l'espoir de maintenir la diversité là où elle se trouve.

BIBLIOGRAPHIE

[BRENZINGER 1992] Brenzinger, Matthias (1992). Lexical retention in language shift: Yaaku/Mukogodo-Maasai and Elmolo/Elmolo-Samburu. In: Brenzinger, Matthias (ed.) *Language Death: Factual and theoretical explorations with special reference to East Africa*. Berlin: Mouton de Gruyter. 213-254.

[BRENZINGER 2003] Brenzinger, Matthias, A. Yamamoto, N. Aikawa, D. Koundioubu, A. Minasyan, A. Dwyer, C. Grinevald, M. Krauss, O. Miyaoka, O. Sakiyama, R. Smeets, & O. Zepeda. (2003). *Language Vitality and Endangerment*. Paris: Unesco Ad Hoc Expert Group Meeting on Endangered Languages. <http://www.unesco.org/culture/en/endangeredlanguages>

[COURLANDER 1996] Courlander, Harold. (1996). *A Treasury of Afro-American Folklore: The Oral Literature, Traditions, Recollections, Legends, Tales, Songs, Religious Beliefs, Customs, and Humor of People of African Descent in the Americas*. New York: Marlowe & Company.

[DIMMENDAAL 1989] Dimmendaal, Gerrit. (1989). On Language Death in Eastern Africa, In: Dorian, Nancy C. (ed.). *Investigating obsolescence: Studies in language contraction*

and death (Studies in the Social and Cultural Foundations of Language, 7). Cambridge: Cambridge University Press. 13-31.

[FISHMAN 1991] Fishman, Joshua. A. (1991). *Reversing language shift*. Clevedon: Multilingual Matters.

[GALLA 2009] Galla, Candace K. (2009). Indigenous Language Revitalization and Technology: From Traditional to Contemporary Domains. In: Reyhner, John and Louise Lockard (eds.). *Indigenous Language Revitalization: Encouragement, Guidance & Lessons Learned*. Flagstaff, AZ: Northern Arizona University. 167-182.

[GIBSON 1996] Gibson, Michael (1996). *El munichi, un idioma que se extingue*. Serie Lingüística Peruana No 42. Pucallpa: Instituto Lingüístico de Verano. [translated by Marlene Ballena Dávila]. <http://www.sil.org/americas/peru/pubs/slp42.pdf>

[HALLET 2007] Hallett, Darcy, Michael J. Chandler and Christopher E. Lalonde. (2007). *Aboriginal language knowledge and youth suicide*. *Cognitive Development*, 22: 3, July-September 2007, 392-399.

[HARRISON 2007] Harrison, K. David. (2007). *When Languages Die: The Extinction of the World's Languages and the Erosion of Human Knowledge*. New York: Oxford University Press.

[LEWIS 2009] Lewis, M. Paul (ed.) (2009). *Ethnologue: Languages of the World, 16th edition*. Dallas: SIL International. <http://www.ethnologue.com>

[LEWIS 2010] Lewis, M. Paul and Gary Simons. (2010). Assessing Endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique LV:2, Special issue on Language Endangerment and Language Death*. 103-120. http://www.lingv.ro/resources/scm_images/RRL-02-2010-Lewis.pdf

[MENSCHING 2000] Mensching, Guido. (2000). *The internet as a rescue tool of endangered languages: Sardinian*. <http://www.gaia.es/multilinguae/pdf/Guido.PDF>

[MICHAEL 2009] Michael, Lev (2009) *National Science Foundation Award Abstract #0941205RAPID: Munique Rapid Documentation Project*. <http://www.nsf.gov/award-search/showAward.do?AwardNumber=0941205>

[MOSELEY 2010] Moseley, Christopher (ed.). (2010). *Atlas of the World's Languages in Danger*, 3rd edn. Paris: Unesco Publishing. <http://www.unesco.org/culture/en/endangeredlanguages/atlas>

[NICOLLE 2012] Nicolle, Steve. *A Grammar of Digo: A Bantu language of Kenya and Tanzania*. Dallas: SIL International and The University of Texas at Arlington.

LE CYBERESPACE DANS L'ÉDUCATION EN LANGUE MATERNELLE

Après avoir redéfini ce qu'est une langue maternelle, nous nous attacherons dans un second temps à évaluer l'usage de la langue maternelle en tant que moyen d'instruction, et à qualifier son enseignement. Puis nous verrons pourquoi et comment le cyberspace peut devenir un terrain d'enseignement des langues maternelles ou en langues maternelles, tant en local qu'à distance.

Article original en français.



MARCEL DIKI-KIDIRI, Centrafricain et Français, est actuellement Consultant en Linguistique Appliquée. Avant sa retraite en 2010, il était Chargé de Recherche au CNRS, dans l'unité mixte Langage, Langues et Cultures d'Afrique Noire (LLACAN : CNRS, INALCO).

MARCEL DIKI-KIDIRI

LE CYBERESPACE
DANS L'EDUCATION
EN LANGUE
MATERNELLE

La question des TIC dans l'éducation en langue maternelle ne peut se poser que si l'on est déjà au clair sur ce qu'on entend par « éducation en langue maternelle ». Dans la plupart des pays du monde, à l'exception des pays en situation de dépendance postcoloniale ou des communautés minoritaires ou minorées, l'éducation se fait dans la langue parlée par la majorité de la population. Dans les communautés monolingues ou très faiblement multilingues, la langue majoritaire est également la langue maternelle de la majorité de ses locuteurs. Dans les sociétés fortement multilingues, la langue majoritaire, lorsqu'elle existe, est bien souvent une seconde, une troisième voire une quatrième langue pour la majorité de ses locuteurs. Il n'en demeure pas moins que dans tous les cas de figure, l'enfant apprend mieux, c'est-à-dire plus de contenus et plus rapidement, lorsqu'il reçoit l'instruction dans la langue qu'il utilise le plus couramment, laquelle est souvent considérée comme sa langue maternelle.

QU'ENTEND-ON PAR « LANGUE MATERNELLE » ?

Le terme même est initialement motivé par l'idée reçue selon laquelle tout enfant apprendrait toujours ses premiers mots sur les genoux de sa mère. La réalité est bien souvent très loin de cette simplification. Dans de nombreuses sociétés patrilinéaires, c'est la langue du père qui est enseignée à l'enfant, que ce soit par sa mère ou non. On aurait pu parler dans ce cas de « langue paternelle » pour décrire cette réalité socioculturelle. Dans les sociétés fortement multilingues, il n'est pas rare que l'enfant apprenne plusieurs langues en même temps dès son premier balbutiement. On parle de « multilinguisme individuel » pour qualifier le fait qu'une même personne parle plusieurs langues, et de « société multilingue » lorsque plusieurs langues sont parlées dans une même communauté vivant sur

un même territoire. La société multilingue est dite diglossique ou polyglossique selon que deux ou plusieurs langues sont parlées par au moins la majorité, si ce n'est la totalité, de la population, avec des fonctions spécialisées pour chaque langue. Mais une société multilingue peut aussi être composée de plusieurs populations locales essentiellement monolingues ayant chacune sa langue maternelle. Enfin, là où une langue véhiculaire est majoritaire, elle devient très vite la première langue que les jeunes générations apprennent. Dès lors, la définition de la langue maternelle en tant que la « première langue qu'apprend un enfant » est en décalage flagrant avec la qualification de « maternelle ». Aussi, les spécialistes de l'éducation préfèrent-ils parler de « langue première » qui est plus juste, plus précise et plus neutre par rapport aux conditions d'apprentissage de la langue. Dans la suite de cet article, il faut comprendre l'expression « langue maternelle » comme un équivalent strict de « langue première ».

QU'ENTEND-ON PAR « ÉDUCATION EN LANGUE MATERNELLE » ?

On parle d'éducation lorsqu'il s'agit de donner aux enfants, aux adolescents et jeunes adultes une formation et une instruction visant à les préparer à assumer pleinement leur place et leur part de responsabilité dans la société en tant qu'adultes et citoyens. Les systèmes d'éducation varient selon les pays, et les communautés humaines en fonction des valeurs essentielles et des savoirs et savoir-faire indispensables que la société choisit de transmettre à sa jeune génération. Lorsqu'il s'agit d'adultes déjà sortis du système éducatif initial, mais qui se retrouvent en situation d'apprentissage, on parle de formation plutôt que d'éducation. La question centrale qui nous intéresse ici est de savoir quelle(s) langue(s) il convient d'utiliser pour enseigner les diverses disciplines au programme dans un système éducatif ou que l'on apprend lors d'une formation professionnelle ?

EN SITUATION MULTILINGUE, DANS QUELLE LANGUE ÉDUIQUER ?

On s'attendrait, a priori, à ce que tout enseignement ou apprentissage utilise la langue que l'apprenant comprend le mieux, laquelle est supposée être sa langue première, sa langue maternelle. Si c'est bien le cas dans

les pays indépendants où les institutions fonctionnent dans la langue de la majorité de la population locale, dans de très nombreux pays fortement multilingues, la langue des institutions n'est pas toujours la langue maternelle de la majorité de la population locale. Les enfants comme les apprenants sont enseignés dans une langue d'instruction qu'ils ne connaissent pas ou ne maîtrisent pas. D'où un effort considérable et disproportionné pour atteindre chaque niveau de savoir dans toutes les disciplines abordées. Est-ce à dire que les pays fortement multilingues sont voués à renoncer au rêve d'une éducation et d'un enseignement en langue maternelle? Grâce à de nombreux travaux de recherche en sciences de l'éducation et des expériences menées dans plusieurs pays fortement multilingues, il est désormais établi qu'il est possible, même dans ce type de pays, de développer divers systèmes éducatifs exploitant judicieusement et positivement le multilinguisme pour offrir aux populations l'accès au savoir dans chacune de leurs langues tout en perfectionnant l'habileté des apprenants à parler plusieurs langues, par l'enseignement de celles-ci. Dans bien des cas, cela nécessite une réforme éducative en profondeur et un effort de longue haleine, mais qui sont, à terme, moins coûteux et bien plus rentables que de demeurer dans le *statu quo* des systèmes inadéquats existants.

La réforme éducative qui vise l'intégration des langues maternelles comme langues d'instruction et comme matières enseignées modifie non seulement les curricula, l'organisation des programmes, le découpage des cycles, etc. Mais elle concerne aussi les méthodes d'enseignement et d'apprentissage qui peuvent intégrer de nouveaux outils pédagogiques, tels que les TIC, et entraîner de nouvelles façons de faire, voire de nouveaux comportements. De plus, l'utilisation d'une langue comme langue d'enseignement présuppose que celle-ci soit suffisamment équipée en terminologie technique de la discipline enseignée pour pouvoir servir à véhiculer les connaissances élaborées dans cette discipline. Le développement systématique de tels vocabulaires spécialisés relève de la terminologie en tant que branche spéciale des sciences du langage et nécessite une mise en œuvre organisée et méthodique par des institutions publiques dédiées (académies, offices, hauts commissariats, délégations, instituts, etc.). Les langues maternelles ainsi équipées devront être enseignées dans un cursus formel et inscrites aux examens d'obtention de diplômes nationaux, pour devenir pleinement attractives. Comme dans tous les domaines d'activités de la vie moderne, les TIC sont également omniprésentes dans le domaine

de l'éducation et notamment de l'enseignement des langues dont elles révolutionnent totalement les méthodes notamment dans le domaine de l'enseignement à distance. Toutefois, l'installation des TIC repose sur de lourdes infrastructures qui ne sont pas toujours présentes dans tous les pays et en particulier dans les zones rurales ou reculées.

QUELLES INFRASTRUCTURES POUR L'UTILISATION DES TIC DANS L'ÉDUCATION EN LANGUE MATERNELLE ?

Deux types d'infrastructures sont à prendre en considération comme pré-requis : celle du système éducatif lui-même et celle des réseaux de communication. Ensuite, il faut des équipements spécifiques pour pouvoir utiliser les TIC dans l'éducation en langue maternelle.

Infrastructures éducatives

La densité et la répartition territoriale des établissements d'enseignement et de formation (toutes les écoles de la maternelle à l'enseignement supérieur, les centres de formation professionnelle, les centres de stages, etc.) le nombre d'apprenants (élèves, étudiants, apprentis, stagiaires, etc.) par établissement et par classe sont des paramètres essentiels qui doivent guider les choix des solutions informatiques favorisant l'accès du plus grand nombre à l'utilisation des TIC. Comme, en général, les villes ont une concentration de population plus élevée que les zones rurales, elles sont aussi plus équipées en infrastructures éducatives. Même très peuplées, les campagnes sont bien plus étendues, ce qui induit une faible densité des infrastructures dans ces zones. De ce fait, les jeunes ruraux ont beaucoup moins de facilité que leurs homologues urbains pour accéder à un établissement scolaire et encore moins à un ordinateur. Ce déséquilibre observable dans les pays riches est encore plus important dans les pays pauvres où le nombre d'élèves par classe est généralement plus que pléthorique.

Infrastructures communicationnelles

Le plus ancien des réseaux de communication encore en usage est bien celui des téléphones filaires qui cède le pas aux ondes radio des réseaux

sans fil. Ceux-ci nécessitent l'implantation de nombreuses antennes relais pour couvrir ne fût-ce qu'une ville à plus forte raison une campagne ou même un continent entier. Les solutions aériennes (satellites géostationnaires, planeurs ultralégers) ne sont pas moins coûteuses et restent moins adaptées à des applications nécessitant des actions rapides et en temps réel (jeux vidéo, chirurgie à distance) mais elles permettent de couvrir de vastes espaces sans d'autres appareils au sol que des récepteurs-émetteurs sans fil. Enfin, la technologie des fibres optiques qui permet le transfert de grandes quantités de données à très haut débit arrive à grands pas, et comme d'habitude, d'abord dans les grandes villes des pays développés avant d'atteindre plus tard les zones rurales et encore plus tard les villes des pays en voie de développement.

Équipements spécifiques

De la même façon qu'une salle de classe doit être équipée de table-bancs ou équivalents, il est nécessaire d'équiper en outils informatiques appropriés tout établissement souhaitant intégrer les TIC dans la panoplie de ses moyens pédagogiques et de gestion scolaire [BASQUE 1998]. Les produits technologiques sont extrêmement variés et à la hauteur de toutes les bourses. Le choix des équipements dépend aussi directement des objectifs et des modes de fonctionnement souhaités. L'ordinateur est le premier outil de base, quelle que soit sa présentation (poste fixe, portable, tablette). Aucun établissement ne peut, cependant, se permettre d'en offrir un à chaque apprenant comme outil de travail. L'aménagement d'une salle informatique où un nombre défini d'ordinateurs montés en réseau local est une solution bien plus gérable. Dans une université où le nombre d'étudiants est habituellement bien plus important que dans un établissement secondaire ou primaire, il peut être utile d'aménager plusieurs salles informatiques regroupées dans un campus numérique. Des logiciels spécifiques permettent de gérer l'ensemble des activités de l'établissement, dont la gestion des cours, des notes, des examens, y compris le flux des informations entre enseignants, étudiants, administratifs [PELGRUM 2004]. Ce dernier point nécessite la mise en place d'un système de messagerie en intranet où chaque membre de l'établissement dispose d'un compte personnel. Enfin, la gestion de la documentation est l'un des domaines de prédilection des TIC. Non seulement elles facilitent la gestion complète d'une bibliothèque locale, mais aussi toutes les relations

entre celle-ci et d'autres fonds documentaires nationaux ou internationaux, permettant ainsi de localiser des ressources disponibles ailleurs et éventuellement d'y accéder.

L'UTILISATION DES TIC POUR L'ENSEIGNEMENT DE LA LANGUE MATERNELLE

Lorsque toutes les conditions d'infrastructures et d'équipements décrits ci-dessus sont réunies, l'utilisation des TIC pour l'enseignement des langues et notamment de la langue maternelle de l'apprenant varie énormément en fonction des approches pédagogiques. Les sciences de l'éducation en ont élaboré de très nombreuses qui toutes ont leurs avantages et leurs insuffisances [BASQUE 2002]. C'est d'ailleurs l'une des raisons principales pour lesquelles elles sont constamment renouvelées. Quelles que soient les théories, aucune ne considère les TIC autrement que comme un dispositif d'aide à l'apprentissage [DEMAIZIÈRE 2007]. Toute la question est de savoir ce que l'on entend par « aide ». L'idée que l'on se fait du processus même de l'apprentissage conditionne ce que l'on considère comme une aide. Si l'on considère, par exemple, que pour atteindre un niveau de compétence acceptable dans la maîtrise d'une langue, il faut réussir à passer un certain nombre de tests avec plusieurs paliers de difficultés étalonnées, on peut penser que cela aiderait l'apprenant de jalonner son parcours avec des indices pour qu'il arrive rapidement à la bonne réponse à chaque test. L'ordinateur est alors utilisé comme un tuteur instructeur et évaluateur [TAYLOR 1980]. Cela va de pair avec une conception de l'apprentissage qui considère l'apprenant comme un cerveau vide à remplir d'informations nouvelles, lesquelles, une fois assimilées, constituent le savoir acquis. Dans cette perspective, on crée des tutoriels censés conduire un « Nul » pas à pas jusqu'à la réussite. Mais de nos jours, au regard des progrès des TIC, conjugués avec ceux des sciences de l'éducation, la pédagogie privilégie l'approche centrée sur l'apprenant [DE VRIES 2001, ANGRIST 2002]. Celui-ci n'est plus un récepteur passif, mais l'acteur de son parcours d'apprentissage. Ses choix et ses actions induisent et construisent la suite de sa progression. Aussi, l'ordinateur devient un outil de production, riche en ressources disponibles [FORCIER 1999]. L'apprenant a toujours des objectifs à atteindre, mais il y va intuitivement en essayant divers outils et en prenant divers chemins selon son tempérament, ses connaissances antérieures, son état psychologique, ses rapports au travail, aux autres, etc. Les autres sont

justement plus présents, car, l'ordinateur n'étant plus qu'un fournisseur de ressources, l'essentiel de l'aide redevient humain. En effet, l'aide provient des pairs à travers plus de travaux collaboratifs rendus possibles par les réseaux des TIC et le télé-enseignement [WILEY 2002] et aussi, des « personnes ressources » que sont les encadreurs, les experts [CAZADE 1999], les intervenants ponctuels, etc. Le métier d'enseignant se fragmente en plusieurs activités spécifiques qui requièrent l'intervention de personnes spécialisées différentes. L'apprenant est invité à « jouer » avec sa panoplie d'outils informatiques afin de les utiliser pour créer et produire en faisant appel à son imagination [JONASSEN 2000, LEBRUN 2002]. C'est dans cet acte de production et de création qu'il avance dans l'appropriation de l'objet principal de ses études, en l'occurrence, sa langue maternelle.

QUELQUES EXEMPLES ÉCLAIRANTS

Au Canada, comme dans la plupart des pays industrialisés, les États ont investi d'importantes ressources dans le réseautage et l'équipement des établissements scolaires primaires, secondaires et supérieurs. La formation initiale des enseignants a été ajustée pour optimiser la capacité des jeunes enseignants à tirer le meilleur parti de ces nouvelles technologies. Malgré cela, de nombreuses études montrent que les enseignants n'en font qu'une utilisation marginale, restreinte à certains types d'activités en classe. Dans de nombreux cas, les usages restent expérimentaux et les résultats deviennent moins significatifs en phase de généralisation. Selon [LAROSE 2010], il faut aller aux États-Unis pour trouver l'expérience la mieux documentée et la plus convaincante :

Une seule étude à échantillon important, méthodologiquement rigoureuse, constate un effet stable du recours aux TIC sur la construction de compétences à l'écrit en langue maternelle. Il s'agit de l'étude évaluative du Maine Middle School Laptop Program où les performances des élèves du secondaire aux tests standardisés après deux ans d'application du programme, tant pour l'organisation des idées que de la grammaire et de la syntaxe, étaient significativement supérieures par rapport à leur niveau de base.

Toutefois, prévient Larose [LAROSE 2010], les conditions pour obtenir de tels résultats sont particulièrement exigeantes. Dans tous les cas où l'on a observé une amélioration significative de la construction des compétences

dans la maîtrise de la langue, tant à l'oral qu'à l'écrit, résultat attribuable à l'utilisation des TIC, on a constaté aussi les quatre conditions suivantes :

- Les élèves ont un accès individuel stable et utilisent un ordinateur en réseau, tant à l'école pour l'ensemble des périodes d'enseignement de la langue, qu'à la maison ;
- Leurs enseignants ont bénéficié d'un haut niveau de soutien tant sur le plan technique que pour l'encadrement des démarches d'intégration pédagogique des TIC ;
- Le curriculum scolaire a aussi été adapté en fonction de la disponibilité d'ordinateurs individuels et des dispositifs pédagogiques particuliers ;
- Des banques exhaustives de situations d'enseignement et d'apprentissage (SEA), adaptées à l'usage massif de l'informatique, ont été développées et régulièrement mises à jour.

Un exemple de ces banques de données est celui du Trésor de la Langue française informatisé (TLFI) qui peut être exploité pour l'enseignement du français, langue maternelle :

Le Trésor de la Langue Française Informatisé permet de faire découvrir aux élèves la facette historique de notre langue. La version informatisée compte 100 000 mots avec leur histoire, 270 000 mots avec leur définition, 430 000 exemples et 500 000 citations. Le moteur de recherche propose trois niveaux de recherche : simple, avancée et complexe. Tous les mots affichés dans un article du TLF peuvent donner lieu à une hypernavigation donnant accès à diverses ressources : Le Trésor de la Langue Française Informatisé, les Dictionnaires de l'Académie (8^e et 9^e éditions), la Base de connaissance lexicale de l'Atilf, Frantext et la Base Historique du Vocabulaire Français.

Pour finir, [LAROSE 2010] propose la conclusion suivante :

D'une façon plus générale, tant nos recherches que l'état de la documentation scientifique suggèrent que l'adoption des TIC en enseignement des langues maternelles ou secondes, tout comme leur intégration dans la pratique d'enseignement de l'ensemble des disciplines scolaires, est tributaire des stratégies de l'intervention éducative adoptées par les enseignants. Dans une perspective d'enseignement relativement traditionnelle, frontale, laissant peu de place à l'initiative de l'élève au regard de la recherche, de l'analyse et de l'intégration de l'information, le recours aux TIC demeure marginal. Au contraire, dans

la perspective d'un enseignement qui accorde une large place à la recherche et à l'intégration de l'information, le rôle des TIC en tant qu'outil de soutien à la construction de connaissances par l'élève apparaît plus évident.

QUELQUES REPÈRES POUR DÉCIDER ET AGIR

En guise de conclusion à cet article, rappelons quelques principes essentiels à prendre en considération dans les décisions à prendre et les choix à faire pour assurer l'utilisation optimale des TIC dans l'éducation en langue maternelle et pour l'enseignement des langues maternelles :

- La langue maternelle est la langue première de l'enfant. C'est elle qui doit servir de langue d'acquisition des premiers savoirs. Bien maîtrisée, elle facilite l'apprentissage d'autres langues, telles que la langue majoritaire lorsque celle-ci est différente ;
- L'utilisation de la langue maternelle comme langue d'enseignement nécessite, d'une part, la mise au point de vocabulaires spécialisés par discipline dans cette langue, et, d'autre part, l'enseignement de la langue ;
- Là où l'éducation n'est pas en langue maternelle, l'intégration de la ou des langue(s) maternelle(s) dans l'éducation nécessite une réforme (assez profonde) de ce système à plusieurs niveaux : formation des enseignants, valorisation des langues maternelles par leur enseignement et leur inscription aux examens officiels ;
- Les TIC apportent de nombreuses formes d'assistance à l'apprentissage. Leur utilisation impose de faire des choix sérieusement étudiés dans différentes approches pédagogiques, car de ces choix dépendent la sélection ou l'élaboration des méthodes et des outils informatiques dont le système éducatif aurait besoin ;
- Pour une utilisation optimale des TIC en éducation, notamment en langue maternelle, les élèves doivent avoir un accès à un ordinateur individuel connecté durant toute la période scolaire. La formation des enseignants ainsi que les curricula doivent être ajustés pour optimiser le recours aux ressources technologiques. Diverses banques de données dans et sur la langue maternelle doivent être développées ou dans tous les cas, rendues accessibles aux élèves, en intranet ou en internet ;

- Il faut toujours garder à l'esprit que plus les ressources informatiques, tant matériels que logiciels, évoluent et se diversifient, plus elles ouvrent de nouvelles possibilités que l'action pédagogique ne se prive pas d'exploiter. L'enseignement à distance, très prisé par les entreprises, a le vent en poupe aussi dans certaines écoles avec la baladodiffusion (utilisant les baladeurs et les téléphones portables) et la vidéoconférence (utilisant les webcams). Les enseignants doivent inventer les cours qui vont avec ces outils devenus courants parmi les jeunes.

BIBLIOGRAPHIE

[ANGRIST 2002] Angrist, J. ; Lavy, V. 2002 « New evidence on classroom computers and pupils learning ». *Economic Journal*, vol. 112, n°482, p. 735-765.

[BASQUE 1998] Basque, J., Rocheleau, J., Winer, L., Michaud, P., Bergeron, G., Paquette, G., Paquin, C. *Un modèle adaptable d'une école informatisée*, Montréal, École informatisée clés en main du Québec inc., 1998.

[BASQUE 2002] Basque, J., Lundgren-Cayrol K. 2002 « Une typologie des typologies des applications des TIC en éducation » *Sciences et techniques éducatives*, n°9 (3-4), p. 263-289.

[BAUMGARTNER 1998] Baumgartner, P., Payr, S. « Learning with the Internet: A typology of application », *Proceedings of ED-MEDIA/ED-TELECOM 98 (World Conference on Educational Multimedia and Hypermedia & World Conference on Educational Telecommunications)*, Charlottesville, AACE, 1998, p. 124-129.

[BONNET 2009] Bonnet, Annick (Coord.) 2009. Elisabeth Brodin, Micheline Maurice, Chirine Anvar, Pernelle Benoit, Séverine Blache, Fiorella Casciato, Concetta Cirocco, Catherine Clément, Stéphanie Favre, Olivier Gisselbrecht, Haydée Maga, Marianne Mavel, Olivier Steffen, Dominique Satgé, Nicole Thiery-Chastel. *Guide SAEL, guide pratique pour la conception, l'animation et l'amélioration des sites d'accompagnement pour les enseignants de langues*. http://www.eurosael.eu/sites/default/files/3/guide_sael_2009_0.pdf.pdf

[CAZADE 1999] Cazade, A. « De l'usage des courbes sonores et autres supports graphiques pour aider l'apprenant en langues ». *Apprentissage des langues et systèmes d'information et de communication (Alsic)*. Vol. 2, n°2, décembre 1999, pp 3 - 32. <http://alsic.revues.org>

[CRABÈRE 2010] Crabère, Béatrice. *Interview. Comment enseigner une langue difficile et minoritaire ?* http://www.cafepedagogique.net/lexpresso/Pages/2010/02/Fourgous_BCrabere.aspx

[DEMAIZIÈRE 2007] Demaizière, Françoise. « Didactique des langues et TIC : les aides à l'apprentissage », *Alsic*, Vol. 10, n°1 | 2007. <http://alsic.revues.org/index220.html>

[DE VRIES 2001] de Vries, E., « Les logiciels d'apprentissage : panoplie ou éventail? », *Revue Française de Pédagogie*, n°137, octobre-décembre 2001, p. 105-116.

[FORCIER 1999] Forcier, R. C., *The computer as an educational tool: Productivity and problem solving*, Prentice-Hall, 1999, 2^e édition, 1999.

[JONASSEN 2000] Jonassen, D. H., *Computers as mindtools for schools: Engaging critical thinking*, Prentice Hall, 2^e édition, 2000.

[LAROSE 2010] LAROSE François, GRENON Vincent, CARIGNAN Isabelle et HAMMAMI Abdelhakim, 2010 « Les TIC en enseignement des langues au Québec : objet obscur d'un désir prescrit ? » *Québec français*, n°159, 2010, p. 71-72. <http://id.erudit.org/iderudit/61597ac>

[LEBRUN 2002] Lebrun, M., *Des technologies pour enseigner et apprendre*, De Bœck, 2^e édition, 2002.

[MUNN 2011] Munn, Yves 2011. *Outils TIC en langues (ESL)* <http://www.reptic.qc.ca/bibliotheque/enquetes-inventaires-compilations/outils-tic-langue-esl.html>

[PELGRUM 2004] PELGRUM, W.J., LAW, N. 2004 *Les TIC et l'éducation dans le monde : tendances, enjeux et perspectives*. Unesco.

[SILVENRAIL 2007] Silvenrail David L. et Gritter Aaron K. 2007 *Maine's middle school laptop program : creating better writers*, Gorham, ME, University of Southern Maine, Maine Education Policy Research Institute.

[WILEY 2001] Wiley, D. A. « Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy », In D.A. Wiley (éd.), *The instructional use of learning objects*, Bloomington, Indiana, Association for Educational Communications of Technology, 2001.

[WILEY 2002] Wiley, D. A., Edwards, E. K. « Online self-organizing social systems: The decentralized future of online learning », *Quarterly Review of Distance Education*, <http://www.opencontent.org/docs/ososs.pdf>

ESPACES NUMÉRIQUES

PARTIE 2

Le cyberspace présente un nombre incalculable de continents invisibles où des milliards de personnes développent ou prolongent des conversations, des relations, des réseaux, des créations, des traductions, des connaissances, des liens sociaux. Quelles passerelles linguistiques peuvent exister pour faciliter ces liens humains en deçà des contraintes techniques et au-delà des barrières culturelles ?

LE MULTILINGUISME DANS LA NORMALISATION DE L'INTERNET

Internet n'existe qu'en fonction du respect de normes, de protocoles, de formats et autres règles collectives nécessaires à l'interconnexion et l'échange. Ces normes peuvent-elles devenir une contrainte limitant l'éventail des possibles ? Quelles sont les normes qui gèrent les langues sur Internet ? Qui les définit, quelles organisations les élaborent ? Sont-elles bien adaptées au multilinguisme ou au contraire sont-elles à la traîne ? Sont-elles un facteur positif pour les langues ?

Article original en français.



STEPHANE BORTZMEYER est ingénieur informaticien, plus particulièrement spécialisé en réseaux TCP/IP. Il travaille pour l'AFNIC et tient un blog où il est question de technique mais parfois aussi de culture ou de politique mais il s'exprime ici à titre personnel. <http://www.bortzmeier.org>

STÉPHANE BORTZMEYER

LE
MULTILINGUISME
DANS LA
NORMATISATION
DE L'INTERNET

Si j'écris du français sur internet, par exemple le simple mot « café », mes lecteurs verront-ils le mot correct ou bien « cafi » ou « caf=E9 » ou encore « cafÃ© » ? Ce phénomène, connu sous le nom japonais de *mojibake*, était encore la règle il y a quinze ans. Ce sont les progrès de la normalisation qui l'ont aujourd'hui considérablement réduit.

Cet article porte sur les différentes normes pratiques qui définissent les protocoles, formats et autres règles que doivent suivre les logiciels qu'on trouve sur internet. Comme dans d'autres secteurs, les normes sont à la fois une bénédiction (sans elles, pas d'internet, puisque pas de communication : imaginez-vous un Web où les sites web seraient visibles uniquement par Firefox ou bien uniquement par Internet Explorer mais pas par les deux ?) et une contrainte puisqu'elles peuvent limiter ce qui est possible. Quelles sont les normes qui s'appliquent au multilinguisme aujourd'hui ? Qui les définit, quelles sdo (*Standards Development Organisation – Organisme de Normalisation*) les élaborent ? Sont-elles bien adaptées au multilinguisme ou au contraire sont-elles à la traîne ? Sont-elles un facteur positif pour les langues ou au contraire une des raisons des problèmes du multilinguisme ?

Je précise que cet article porte bien sur les *normes*, pas sur les *mises en œuvre* de celles-ci. Une norme n'est pas tout. Si elle n'est appliquée par aucun programme, ou bien si aucun utilisateur ne s'en sert, elle n'a pas d'utilité, elle sera une « oubliée numérique ». On verra que, une fois la norme terminée, il reste du travail pour les programmeurs (pour l'implémenter), pour les administrateurs systèmes et réseaux (pour la déployer) et pour les utilisateurs (pour... l'utiliser).

Je vais me permettre un petit rappel sur le multilinguisme et son traitement par l'informatique. Internet étant traditionnellement basé sur l'écrit,

il faut bien faire la différence entre *langue* et *écriture*. Une écriture peut servir à plusieurs langues (c'est le cas de l'alphabet latin) et une langue peut utiliser plusieurs écritures (c'est le cas de l'azéri¹ ou du tamasheq²). Il est donc important de bien distinguer quand on a besoin de gérer plusieurs langues et quand il faut gérer plusieurs écritures.

D'autre part, les langues humaines n'ont pas été conçues scientifiquement : elles sont bizarres, ont plein de règles incompréhensibles et beaucoup d'idiosyncrasies. Si on pouvait refaire les langues à partir de zéro, dans le seul but de faciliter leur informatisation, on simplifierait beaucoup le problème du multilinguisme. Mais ce n'est évidemment pas la politique des organismes de normalisation, qui considèrent tous que l'état actuel des langues et des écritures est donné : il n'est pas envisageable de le modifier. Ainsi, lorsque certaines personnes disent que la norme Unicode, par exemple, est complexe, elles se trompent de cible. Ce sont les écritures, ces créations humaines, qui sont complexes et Unicode ne fait que refléter ce monde réel³.

Autre obstacle au multilinguisme : le manque de connaissances scientifiques sur certaines langues. La fracture numérique se manifeste également là, certaines langues ne sont pas représentées sur internet parce qu'elles n'ont pas encore fait l'objet d'une modélisation rigoureuse.

Enfin, une difficulté supplémentaire vient de la grande sensibilité que posent les questions de langues et d'écriture. Toute divergence technique est vite perçue comme une atteinte à la sensibilité d'une nation ou d'un peuple, ce qui rend souvent les discussions peu sereines.

Autre rappel qui peut être nécessaire : le fonctionnement de l'internet lui-même. Il est important de comprendre qu'internet n'a pas de Direction Centrale qui pourrait par exemple donner des consignes comme « Au 15 janvier 2011, tous les logiciels de courrier doivent accepter des adresses en Unicode ». Le déploiement de toute technologie va dépendre au contraire des décisions d'un certain nombre d'acteurs (parfois un très grand nombre) et il faudra donc les convaincre. De ce fait, et par suite des énormes investissements qui ont été faits depuis trente ans, le poids

1 Écrit aujourd'hui avec les alphabets latin ou arabe et qui l'a longtemps été avec le cyrillique.

2 Écrit en alphabet latin mais surtout en tfinagh.

3 La complexité – bien réelle – d'Unicode provient aussi de contraintes techniques, comme la nécessité d'une compatibilité avec les codages préexistants.

de l'historique est élevé : on ne peut pas tout jeter pour tenter de refaire internet « en mieux » pas plus qu'on ne peut raser une ville pour faire un meilleur urbanisme.

LES NORMES QUI PEUVENT INFLUER SUR LE MULTILINGUISME

En quoi une norme, document en général plutôt technique et aride, peut-elle favoriser ou gêner le multilinguisme ? Prenons quelques exemples de l'importance des normes. Un des cas les plus connus est celui des *jeux de caractères*. Pour écrire une langue avec un ordinateur, il faut pouvoir représenter chaque caractère du texte par un nombre, les seuls objets que puissent manier des programmes. Il faut donc établir une liste des caractères (c'est déjà un effort de formalisation qui n'est pas évident, et qui n'est pas encore fait pour toutes les écritures) et attribuer à chacun un nombre. Si deux personnes qui écrivent en tamoul se servent de deux jeux de caractères différents⁴, les messages de l'un seront incompréhensibles à l'autre. Une norme pour les jeux de caractères est donc nécessaire. Mais encore faut-il qu'elle couvre toutes les écritures du monde ! L'une des premières normes de caractères, ASCII (*American Standard Code for Information Interchange*), conçue en 1969 pour l'usage de l'anglais, ne couvre ainsi pas les caractères utilisant des signes diacritiques (comme é ou ç) ni, a fortiori, les caractères des écritures arabe, devanagari ou hangul. Après ASCII, plusieurs normes incompatibles⁵ ont été conçues sans qu'aucune ne couvre toutes les écritures du monde, avant l'apparition d'Unicode, présenté plus loin.

Une fois qu'il existe une norme pour les caractères, encore faut-il que les différents formats permettent son utilisation. Ainsi, à ses débuts, le courrier électronique sur internet ne permettait que l'utilisation de l'ASCII. Les utilisateurs d'une autre écriture devaient, soit renoncer à certains caractères pour ceux qui utilisaient l'alphabet latin⁶, soit mettre au point un système de translittération de leur écriture en ASCII⁷. Il faudra

4 Mettons que l'un écrive en Unicode et l'autre en ISCII.

5 Et ne couvrant que des « grandes » écritures. Les « petites » sont celles qui ont le plus bénéficié d'Unicode.

6 Par exemple, en français, on peut remplacer les caractères composés par leur équivalent ASCII et le texte reste relativement lisible dans la plupart des cas.

7 Comme le rōmaji pour le japonais qui, en pratique, n'a guère connu de succès.

de longues discussions au sein de l'IETF (*Internet Engineering Task Force*), l'organisme qui propose les normes pour internet, pour que soit établi un document de normalisation (une RFC: *Request For Comments*) pour étendre l'usage des jeux de caractères. Les autres jeux de caractères ne seront acceptés dans le courrier qu'en 1992, avec la sortie du RFC 1341.

Autre cas où les normes et le multilinguisme ont agité beaucoup d'électrons: les identificateurs. Tous les protocoles internet définissent des *identificateurs* légaux et d'autres qui ne le sont pas. Comme ces identificateurs sont très visibles (on les trouve sur les publicités, sur les cartes de visite, etc.), la gêne peut être considérable. Ainsi, dans les adresses de courrier électronique, comme `stephane@bortzmeyer.org`, seul l'ASCII était autorisé jusqu'à la sortie du RFC 4952. Désormais (mais attention, ce RFC est encore considéré comme expérimental, seul son successeur, en cours de mise au point, sera une véritable norme), on peut avoir des adresses comme `stéphane@bortzmeyer.org`

Autre cas qui a beaucoup fait parler de lui⁸, les noms de domaine en Unicode, les IDN (*International Domain Name*). Pour diverses raisons⁹, les noms étaient traditionnellement restreints à US-ASCII seul. La sortie de la norme IDN en 2003, dans le RFC 3490, a marqué le début des noms en Unicode, qui sont devenus banals aujourd'hui. Contrairement à d'autres normes portant sur l'internationalisation, ils ont en effet été mis en œuvre très vite dans les logiciels et déployés dans plusieurs registres de noms de domaine.

UN PREMIER EXEMPLE DE NORME : UNICODE

Historiquement, toutes les normes de jeux de caractères étaient limitées à une écriture, ou à un ensemble d'écritures proches¹⁰. Une des conséquences de cette Babel des jeux de caractères est qu'il était très difficile d'écrire un texte comportant plusieurs écritures (un cours de hindî écrit en espagnol, par exemple). Certains jeux de caractères comportaient un moyen d'inclure de l'ASCII mais pas, en général, les autres caractères

8 Et pas forcément de manière justifiée.

9 Parmi lesquelles ne figure pas une « limite du DNS ». Ce dernier acceptait tous les caractères depuis le début. Mais il est utilisé dans de nombreux contextes, dont certains auraient posé un problème avec des noms en Unicode.

10 Par exemple, ISCII couvre toutes les écritures qui ont un statut officiel en Inde et qui sont dérivées de l'écriture brāhmī.

de l'alphabet latin, sans même parler des autres écritures. D'autre part, gérer une collection de textes écrits avec différentes écritures (même si chaque texte n'en comportait qu'une) était difficile. Par exemple, pour un serveur web, il ne serait pas possible, sans Unicode, de configurer un paramètre `Charset` global, qui indiquerait le codage des caractères pour tout le site, même multilingue.

Unicode a changé tout cela : Unicode¹¹ est un jeu de caractères qui englobe toutes les écritures du monde¹². Que contient la norme Unicode ? D'abord, Unicode est une liste de caractères. Compiler une telle liste ne semble rien, mais, en fait, c'est une tâche difficile. Si certaines écritures sont très normalisées, par exemple, et qu'il suffit donc de reprendre cette norme. Dans d'autres cas, il n'existe pas de liste officielle. Les auteurs de la norme doivent donc se poser des questions comme « Le ß allemand majuscule existe-t-il ? »¹³ ou bien « Les caractères japonais et chinois sont-ils les mêmes ? »¹⁴. Une fois qu'on a donné une réponse à toutes ces questions, on peut publier la liste¹⁵. Elle comprend actuellement 109 449 caractères, des plus banals comme le « petit a » latin aux plus étonnants comme le « lever de soleil sur les montagnes »¹⁶.

Une fois la liste établie, Unicode donne à chaque caractère un numéro unique, qui facilite les discussions : lorsque deux caractères se ressemblent visuellement, ou bien lorsque les polices nécessaires ne sont pas installées, ce numéro permet un échange sans ambiguïté. Pour donner un exemple, les deux caractères cités plus haut, *a* et *lever-de-soleil-sur-les-montagnes* ont respectivement les numéros U+0061 et U+1F304¹⁷.

Mais ces numéros servent aussi de base à l'encodage ultérieur de ces caractères. En effet, il va falloir aussi représenter ces caractères par une suite de bits dans les fichiers ou sur le réseau. Il existe plusieurs méthodes

11 <http://unicode.org>

12 En novembre 2010, la version actuelle d'Unicode étant la 6.0, certaines écritures sont encore absentes mais ce sont presque toutes des écritures mortes, dont le manque ne touche que les chercheurs.

13 Non jusqu'à Unicode 5.2, oui après.

14 La réponse a été Oui, cela se nomme l'« unification Han » et c'est certainement la décision la plus controversée d'Unicode.

15 Un des points importants d'Unicode est que, non seulement le texte de la norme, mais également les données – comme la liste des caractères – sont distribués au public.

16 Voir le caractère sur le site Unicode : <http://www.fileformat.info/info/unicode/char/1f304/index.htm> 

17 Les numéros sont conventionnellement précédés de U+ et écrits en hexadécimal.

pour cela, connues par des noms comme UTF-8 ou UTF-32, et qui toutes partent du numéro pour le représenter d'une manière adaptée à certains usages. En pratique, ce dernier point ne concerne que les informaticiens.

Tout aussi technique mais peut-être plus nécessaire à comprendre, les concepts de canonicalisation : il existe plusieurs façons de représenter le même caractère visuel en Unicode. Par exemple, le é de mon prénom peut être représenté par U+00E9 (e avec accent aigu) ou bien par U+0065 U+0301 (e suivi d'un accent aigu combinant). Des opérations courantes en informatique comme la comparaison (imaginez un utilisateur dont l'identifiant de connexion serait le prénom...) échoueraient si on les appliquait naïvement entre caractères Unicode. Il est donc nécessaire de canonicaliser les chaînes de caractères, de les réduire à une forme cano-nique. La plus courante des normes de canonicalisation, sur l'Internet (cf. RFC 5198) est dite NFC et, dans le cas présenté plus haut, réduirait tous les é à la forme U+00E9.

Au fait, qui rédige et maintient cette norme ? Le consortium Unicode est un regroupement de plusieurs organisations, notamment des gros industriels de l'informatique (Google, Apple, IBM, Microsoft, etc.). Récemment, des organisations sans but lucratif ont commencé à s'intéresser à ce problème et ont rejoint le consortium. S'il existe une très intéressante liste de discussion publique, unicode@unicode.org, l'essentiel du travail se fait en privé, seuls les résultats étant publics.

La meilleure lecture en français, pour ceux qui veulent approfondir Unicode, est le livre de Patrick Andries [ANDRIES 2008].

UN EXEMPLE DE SDO : L'IETF

Faisons un petit détour par une SDO (*Standards Developing / Development Organization*) particulièrement importante, l'IETF (*The Internet Engineering Task Force*). Cette organisation est entre autres responsable des normes sur le courrier électronique, sur le protocole de messagerie instantanée XMPP, sur le protocole HTTP, sur le DNS, etc. L'une des particularités de l'IETF est son extrême ouverture : il n'y a pas d'adhésion formelle, donc pas de cotisation. N'importe qui, individu ou entreprise, peut participer. S'il ne peut pas se déplacer aux réunions physiques (qui sont, elles, assez chères), ce n'est pas forcément dramatique, certains

groupes de travail IETF ne se sont jamais réunis en présentiel¹⁸. Même si le groupe se réunit physiquement, l'essentiel du travail est fait en ligne, via des listes de diffusion publiques (et archivées publiquement) et des documents de travail également publics. WikiLeaks n'aurait pas grand-chose à faire pour assurer la transparence de l'IETF :-)

Quelle est la politique de l'IETF en matière de multilinguisme? Celle-ci est exposée dans le RFC 2277. En deux mots, l'IETF sépare les *éléments de protocole*, internes au fonctionnement du protocole, du texte qui, lui, est montré aux utilisateurs. Les premiers sont visibles uniquement des programmeurs. Ainsi, un navigateur web qui demande la ressource `/faq.html` à un serveur envoie la commande `GET /faq.html`. Le verbe `GET` est certes tiré de l'anglais mais ce n'est pas réellement un mot anglais, plutôt un élément du dialogue `HTTP`. L'utilisateur ordinaire ne le voit jamais et il n'y a donc pas de raison de le traduire. En revanche, le texte de la page web ainsi récupérée va être vu par un utilisateur. Là, le RFC 2277 pose comme principe qu'il doit pouvoir être encodé dans n'importe quel jeu de caractères et ne doit surtout pas être restreint à `ASCII`.

Il s'agit là d'excellents principes mais, évidemment, la réalité est plus complexe. Deux cas ne sont pas directement traités par ce RFC, celui, très sensible, des *identificateurs* (comme les noms de domaine ou bien les adresses de courrier électronique cités plus haut), qui sont à la fois éléments de protocole et textes lus par les utilisateurs. Une bonne partie des polémiques autour des `IDN` (*Internationalised Domain Names*), par exemple, vient du choc des deux points de vue, ceux qui considèrent un nom de domaine comme un identificateur formel, dénué de sémantique (et qui peut donc être écrit dans un alphabet étranger à l'utilisateur) et ceux qui le considèrent comme un marqueur d'identité, qui doit être lisible.

À noter que le `w3c`, organisation de normalisation des technologies du web, qui a un fonctionnement relativement proche de l'IETF, a une politique analogue¹⁹.

18 Comme le groupe de travail `LTRU`, qui a créé les étiquettes de langues mentionnées plus loin.

19 <http://www.w3.org/International/getting-started>

DEUXIÈME EXEMPLE DE NORME : LE COURRIER ÉLECTRONIQUE

Le courrier électronique est une des applications les moins visibles et pourtant les plus utilisées de l'internet. Malgré certains pronostics comme quoi il serait appelé à disparaître devant la messagerie instantanée, ou devant les outils de communication fermés et contrôlés de services comme Facebook, des millions de messages continuent de s'échanger chaque jour. Comment le courrier gère-t-il l'internationalisation ? Il y a deux problèmes distincts, le contenu des messages et les adresses.

Autrefois, le seul contenu accepté était du texte brut, exclusivement en US-ASCII. Cela a changé en 1992 avec la parution de la norme MIME (*Multipurpose Internet Mail Extensions*). Celle-ci permettait de mettre dans un message des instructions de mise en page pour le texte, mais aussi du son, des images, etc. Un autre aspect de MIME avait été moins remarqué à l'époque : le jeu de caractères du texte n'était plus obligatoirement de l'US-ASCII : tout jeu de caractères était accepté, à condition d'être proprement indiqué... et que le logiciel du récepteur puisse l'utiliser. Depuis, cette époque, on peut dire que les normes du courrier permettent d'écrire un message en n'importe quelle langue, mais il a fallu assez longtemps pour que tous les auteurs de logiciels s'adaptent. On voit ici qu'établir des normes n'est que le premier pilier d'une politique linguistique dans le cyberspace. Les incitations, en direction des utilisateurs, ou des programmeurs, pour les faire appliquer participent aussi d'une décision avisée.

Jusqu'à très récemment, il restait un manque : les adresses de courrier elles-mêmes n'étaient pas internationalisées. Pas question de mettre sur sa carte de visite `stéphane@coopération.com`. Cette limite a sauté en 2008, avec les RFC expérimentaux déclinant le RFC 4952. Peu déployée pour l'instant, cette possibilité devrait se répandre avec son prochain accès au statut de norme. On imagine aisément l'intérêt pour les écritures non-latines, pour lesquelles la translittération des noms ne serait plus nécessaire.

TROISIÈME EXEMPLE DE NORME : LES ÉTIQUETTES DE LANGUE

Beaucoup moins connues, car moins visibles, que MIME ou que les IDN, les étiquettes de langues (*language tags*) sont de courts identificateurs qui

servent à indiquer la langue d'un document. Elles sont indispensables aux bibliothécaires et documentalistes, aux linguistes qui échangent leurs documents, mais aussi aux auteurs de sites web lorsqu'ils veulent indiquer la langue du document, à des fins de présentation (les règles typographiques ne sont pas les mêmes pour toutes les langues), de recherche (faciliter la tâche du moteur de recherche lorsqu'on lui demande « uniquement les documents en portugais »). Un format comme XML permet ainsi d'indiquer la langue du document, évitant au logiciel d'édition de devoir deviner, opération complexe et pas toujours sûre. Si aujourd'hui les étiquettes de langue restent malheureusement peu utilisées sur le Web (aussi bien par les auteurs de pages que par les moteurs), elles sont par contre très présentes dans les grands catalogues documentaires.

Normalisées dans le RFC 5646, les étiquettes de langue²⁰ permettent d'indiquer la langue mais aussi l'écriture utilisée, les variantes nationales ou régionales, etc. Ainsi, si la simple étiquette `el` désigne simplement le grec moderne, sans autre précision, l'étiquette plus complexe `guc-Latn-HK` indique le cantonais utilisé à Hong-Kong, écrit en alphabet latin.

Normaliser de tels identificateurs n'a pas été un chemin de roses. Tout ce qui touche aux langues est extrêmement sensible, et le choix de considérer telle idiome comme une langue ou un dialecte n'est pas neutre et peut susciter colère et incompréhension. C'est en partie pour limiter ce risque que les étiquettes de langue s'appuient, autant que possible, sur d'autres normes comme ISO 639 pour les langues. Le RFC 5646 apporte, par rapport à ces normes, la libre disponibilité de la norme, la possibilité de combinaison (comme l'exemple cité plus haut) et la stabilité (contrairement aux identificateurs ISO, une étiquette reste toujours valable, même si l'identificateur est retiré ou réaffecté par l'ISO).

HISTOIRE

Si la situation du multilinguisme sur internet aujourd'hui est assez bonne (quasi-parfaite pour ce qui est de la normalisation, moins idéale pour la mise en œuvre et le déploiement), cela n'a pas toujours été le cas. Tous les utilisateurs au-delà d'un certain âge se souviennent de l'époque où le simple fait d'envoyer un message avec un caractère composé nécessitait de bonnes compétences en informatique et la lecture de beaucoup

²⁰ <http://www.langtag.net>

de documentations. L'auteur se souvient du premier logiciel de lecture MIME²¹, et a gardé un souvenir ému de la longue lutte des années 1990 pour que deux français puissent s'envoyer des messages en français correct. Félicitons donc a posteriori les membres du groupe GERET²², qui avait fait un énorme travail de sensibilisation et de formation.

Cette longue lutte a laissé des traces, notamment la légende urbaine tenace comme quoi « Internet n'accepte pas les accents », ce qui conduit certains francophones à s'auto-limiter, encore aujourd'hui, et à n'utiliser que l'ASCII. Certes, leur attitude est justifiée par le fait qu'on ne peut pas aujourd'hui garantir 100 % de succès mais, malheureusement, elle contribue à bloquer le progrès : en 2011, les utilisateurs ne devraient plus tolérer un système qui interdit d'utiliser tous les caractères de leur langue !

Et l'avenir, que nous réserve-t-il ? Début 2011, le travail de normalisation est fini à 95 %²³ et le problème est désormais surtout de programmation, de déploiement et de contenu. Du travail concret attend ceux et celles qui veulent vraiment aider au multilinguisme sur internet !

GLOSSAIRE

American Standard Code for Information Interchange (ASCII)

Un ancien (mais toujours très utilisé) jeu de caractères normalisé aux États-Unis et qui n'a donc que les caractères nécessaires pour l'anglais.

Vu qu'il était l'un des premiers, et vue sa naissance au pays de l'informatique, il a longtemps servi de base à beaucoup de protocoles réseau.

Domain Name System (DNS)

Ce terme désigne à la fois le système de noms de domaine, système arborescent permettant de créer des identificateurs (`cooptel.qc.ca` ou `véliplanchiste.com`) et le protocole permettant, à partir d'un tel nom, de récupérer des informations comme l'adresse IP, le nom du serveur de messagerie, etc.

Internationalized Domain Names (IDN)

Le terme IDN désigne les noms de domaine exprimés en Unicode comme, par exemple, سنوت.تيرح²⁴. On utilise parfois le sigle IDNA (*Internationalized Domain Names in*

21 Il était nommé metamail et c'était un programme dont même l'informaticien le plus dinosaurien ne voudrait pas aujourd'hui.

22 Groupe d'Exploitation des Réseaux Ethernet TCP/IP. C'était un groupe de travail, dont le but était de constituer un lieu d'échange d'expériences pour des ingénieurs qui exploitent des réseaux à dominance Ethernet et TCP/IP.

23 Les deux manques principaux sont le texte Unicode dans FTP - *File Transfer Protocol* - et le passage des adresses de courrier en Unicode au statut de norme complète.

24 Qui utilise le domaine de tête national de la Tunisie.

Applications) pour la technique spécifique employée actuellement, qui passe par une conversion locale en ASCII, avant l'envoi au DNS.

Internet Engineering Task Force (IETF)

La principale organisation de normalisation de l'internet, notamment chargée des couches 3 (routage) à 7 (applications). L'IETF se signale par sa grande ouverture, ses débats et ses normes (les fameux RFC) étant publics.

Requests for comments (RFC)

Les *Requests For Comments* (RFC), littéralement « demande de commentaires », sont une série numérotée de documents officiels décrivant les aspects techniques d'internet, ou de différent matériel informatique (routeurs, serveur DHCP). À noter que tous les RFC n'ont pas le statut de normes, certains sont qualifiés de « pour information seulement » et d'autres d'« expérimentaux ».

Standard Development Organisation (SDO)

Une organisation, en général sans but lucratif, qui développe et maintient des normes. Ce terme est en général réservé à celles relativement ouvertes (comme l'IETF, l'UIT ou le W3C), au détriment d'organisations représentant plutôt un cartel de certains industriels.

World Wide Web Consortium (W3C)

Organisation de normalisation des formats liés au Web, comme HTML (format des pages web), XML (format pour des données structurées) ou CSS (mise en page des pages web). <http://www.w3.org/>

BIBLIOGRAPHIE

[ANDRIES 2008] Patrick Andries. *Unicode en pratique*. 2008. Dunod.

[UNICODE STANDARD] The Unicode Consortium. *The Unicode Standard, Version 6.0.0*. 2010. The Unicode Consortium.

[GILLAM 2002] Richard Gillam, *Unicode Demystified: A Practical Programmer's Guide to the Encoding Standard*, Addison-Wesley, 2002.

[KORPELA 2006] Jukka K. Korpela, *Unicode Explained*, O'Reilly, 2006.

RFC

[RFC 1341] N. Borenstein. N. Freed. *MIME (Multipurpose Internet Mail Extensions): Mechanisms for Specifying and Describing the Format of Internet Message Bodies*. 1992.

[RFC 2277] H.T. Alvestrand. *IETF Policy on Character Sets and Languages*. 1998.

[RFC 3490] P. Faltstrom. P. Hoffman. A. Costello. *Internationalizing Domain Names in Applications (IDNA)*. 2003.

[RFC 4952] J. Klensin. Y. Ko. *Overview and Framework for Internationalized Email*. 2007.

[RFC 5198] J. Klensin. M. Padlipsky. *Unicode Format for Network Interchange*. 2008.

[RFC 5646] A. Phillips. M. Davis. *Tags for Identifying Languages*. 2009.

MIKAMI YOSHIKI
& SHIGEAKI KODAMA

MESURER LA DIVERSITÉ LINGUISTIQUE SUR LE WEB

La question de la localisation dans la société de l'information a éveillé de la curiosité, mais également beaucoup d'inquiétude chez un grand nombre de chercheurs. Les auteurs de cet article ont créé en 2003 le *Language Observatory Project* afin de mesurer le degré d'utilisation de chaque langue. Si chacun s'accorde sur la nécessité d'une telle évaluation, la méthodologie de l'observatoire et les résultats de ses recherches méritent attention pour qui veut comprendre l'état de la diversité linguistique dans le monde numérique.

Article original en anglais.
Traduit par Laurent Vannini.



MIKAMI YOSHIKI est le directeur de l'Observatoire des langues (*Language Observatory Project*) créé à l'Université de Technologie de Nagaoka (Japon) en 2003.

SHIGEAKI KODAMA a rejoint l'équipe de recherche en 2006. Le laboratoire étudie la diversité linguistique dans le cyberspace, et évalue régulièrement son évolution.

Avec la collaboration de **CHEW YEW CHOONG, PANN YU MON, OHNMAR HTUN, TIN HTAY HLAING, KATSUKO T. NAKAHIRA, YOKO MITSUNAGA**

MIKAMI YOSHIKI
& SHIGEKI KODAMA

MESURER
LA DIVERSITÉ
LINGUISTIQUE
SUR LE WEB

Les développements rapides qui affectent les technologies de l'information modifient radicalement la communication à l'échelle mondiale, multipliant ses formes et zones d'action. Néanmoins, ces nouvelles technologies ne bénéficient pas de la même manière à toutes les communautés linguistiques, laissant dès lors émerger une possible « fracture linguistique numérique. » Prenons pour référence un épisode de l'histoire de la révolution de l'imprimerie. En 1608, Thomas Stephens, un moine jésuite installé sur la cote sud-ouest de l'Inde, écrit à ses supérieurs à Rome :

« Avant de conclure cette lettre, je souhaiterais attirer l'attention de Votre Sainteté sur le fait que depuis de nombreuses années je désire ardemment voir dans cette province des livres imprimés dans les langues et alphabets locaux, comme on en trouve à Malabar pour le plus grand bénéfice de la communauté chrétienne. Et cela n'a pas pu se faire pour deux raisons ; la première étant qu'il semblait impossible de fabriquer tous les moules nécessaires pour leurs six cents caractères, à comparer aux vingt-quatre utilisés en Europe. » [PRIOL-KAR 1958]

Au moment où le moine écrit cette lettre, plus de cent cinquante ans se sont écoulés depuis l'innovation de Gutenberg, mais la nouvelle technologie d'impression n'atteindrait pas sa paroiisse avant le XIX^e siècle. Comme il l'indiquait lui-même, le principal obstacle était la difficulté à introduire les technologies de l'imprimerie dans les langues locales. Dans la terminologie contemporaine on traiterait de cet obstacle sous le terme de localisation ou encore régionalisation. La difficulté de réaliser les moules en métal d'un grand nombre caractères d'imprimerie prend une tout autre forme à l'ère des ordinateurs et de l'internet.

La question des problématiques de localisation dans une société de l'information a suscité intérêt et inquiétude chez un grand nombre de chercheurs, dont nous sommes. En 2003, nous avons démarré le *Language Observatory Project* (Projet d'Observatoire Linguistique) avec comme ambition de mesurer l'utilisation respective de chaque langue dans le cyberspace.

La première partie de cet article invoque les raisons justifiant de telles mesures. Nous présenterons ensuite le *Language Observatory Project*, puis nous décrirons des résultats récents issus de nos observations.

DE LA NÉCESSITÉ DE MESURER

La localisation plus que jamais d'actualité

Les moules permettant la fonte des caractères dans l'imprimerie correspondent aujourd'hui aux jeux de caractères codés des technologies informatiques. Nous disposons désormais d'une norme internationale de codage des caractères pour l'échange d'information, l'ISO/IEC 10646 définissant un jeu universel de caractères abstraits codés sur plusieurs octets, l'Universal Character Set (UCS), ou Unicode¹. Comme son nom l'indique, elle regroupe un univers entier de codes de caractères allant des anciens systèmes d'écritures cunéiformes et hiéroglyphiques égyptiens, à des écritures minoritaires telles celles utilisées dans les régions montagneuses reculées d'Asie du Sud-Est par les locuteurs des langues thai-kadai.

Toutefois de nombreuses difficultés persistent dans le traitement du langage. La plus importante d'entre elles étant que l'UCS, contrairement à ce que son nom indique, n'inclut pas l'intégralité des jeux de caractères utilisés par l'humanité ; selon notre étude, un grand nombre de locuteurs de langues singulières rencontrent les mêmes obstacles que le moine jésuite évoqué plus haut a rencontré dans l'Inde du XVI^e siècle.

La langue mongole, par exemple, est écrite soit en alphabet cyrillique, soit dans son propre alphabet historique et traditionnel, pour lequel au moins

1 Unicode est une norme créée par Le Consortium Unicode. Mais son développement et ses modifications sont parfaitement synchronisés avec la norme officielle ISO/IEC 10646. Ces deux normes peuvent être envisagées comme une seule et même norme. Voir dans cet ouvrage : Stéphane Bortzmeyer, *Normes de l'internet et multilinguisme*.

huit types de codes de caractères différents ont été identifiés². Aucune standardisation n'existe pour ces jeux de caractères, ce qui est source d'inconsistance et parfois même de transcriptions textuelles erronées d'un ordinateur à un autre. Par conséquent, certaines pages web mongoles sont composées par des fichiers images, qui prennent plus de temps à charger.

Les pages web créées en Inde connaissent les mêmes déboires. Sur des sites propriétaires de journaux indiens, des fontes spécifiques à l'écriture hindi sont souvent utilisées et certains d'entre eux diffusent leurs informations sous la forme de fichiers images. Ces barrières technologiques empêchent l'interopérabilité de l'information et conduisent à une fracture linguistique numérique. Notre recherche montre que si l'utilisation des caractères hindi de l'UCS hindi se répand, de nombreuses pages web dépendent toujours de l'affichage de fichiers images ou de l'utilisation de polices de caractères propriétaires.

De tels challenges technologiques maintiennent non seulement l'écart entre les langues mais également entre les écritures. La motivation des auteurs de ce travail émerge de ce constat. Lorsque fut créé le *Language Observatory Project*, l'un de ses fondateurs rédigea le communiqué suivant :

« Mes récents travaux à partir de données statistiques fournies par l'UIT et l'Unesco donnent un premier aperçu de la fracture numérique mondiale « parmi les écritures ». Les utilisateurs de l'alphabet latin, qui représentent 39 % de la population mondiale, consomment 72 % de la production imprimée et écrite sur papier dans le monde, et profitent de 84 % de l'ensemble des connexions à l'internet. Les utilisateurs de sinogrammes – caractères chinois, en Chine, au Japon et en Corée, soit 22 % de la population mondiale, consomment 23 % de la production papier et comptent pour 13 % des accès internet. Les utilisateurs de langue arabe, 9 % de la population globale, consomment 0.5 % de la production papier et représentent 1.2 % des accès à l'internet. Les utilisateurs de l'alphabet cyrillique, 5 % de la population, totalisent 1.1 % des écrits papier et 1.6 % des accès à l'internet. Qu'en est-il alors des utilisateurs d'écritures indo-aryennes ? Si l'on y adjoint toutes les écritures d'origine brahmiques utilisés massivement dans l'Asie du Sud – myanmar, thaï, lao, khmer, etc., les utilisateurs

2 En plus de l'UCS/Unicode, BeiDaFangZheng, GB18030, GB8045, Menksoft, Sayinbilig, Boljoo and SUDAR. La plupart de ces systèmes d'encodage sont des codes locaux et propriétaires utilisés par un nombre limité de personnes.

d'écritures indiques représentent 22 % de la population mondiale, consomment 2.2 % des écrits papier et n'ont que 0.3 % des connexions à l'internet. [MIKAMI 2002].

Le *Language Observatory Project* a été créé pour identifier et combler ces écarts afin d'assurer l'égalité et la diversité sur la toile.

La domination de l'anglais sur le Web

La seconde raison expliquant la fracture numérique linguistique est la domination exercée par l'anglais sur la Toile, qui peut refléter également l'aspect économique de l'évolution du Web. Ce sujet fut abordé pour la première fois en 1995 au Sommet de la Francophonie de Cotonou. Durant cette rencontre, la présence de l'anglais sur la Toile fut publiquement évaluée comme ayant dépassé la barre des 90 %. FUNDERES (*Fundación Redes y Desarrollo*), en réaction à ce chiffre, tenta d'obtenir une mesure précise de la présence de plusieurs langues dont l'anglais sur la toile [PIMIENTA ET AL. 2010].

Notre recherche montre que dans les domaine nationaux de premier niveau (ccTLD) tels que les sites en .sn pour le Sénégal par exemple, l'anglais continue à dominer, plus de 10 ans après ce sommet. De 2006 à 2009, nous avons conduit des enquêtes annuelles de présence des langues sur la Toile. Une comparaison directe est impossible puisque le nombre de pages collectées diffère pour chaque étude (nous essayons actuellement d'identifier une méthodologie pour normaliser l'impact de la disparité de la taille des échantillons, en utilisant l'analyse de la variance). Mais nous pouvons néanmoins faire l'observation que dans toutes les enquêtes, l'anglais restait la langue la plus utilisée dans les domaines asiatiques et africains.

Les mesures effectuées en 2010 révélèrent que l'anglais était utilisé dans 82.7 % des pages collectées depuis les domaines de premier niveau africains; Le français était la seconde langue la plus usitée avec 5.5 % des pages. Pour les domaines asiatiques, l'anglais était également à la première place, mais dans des proportions plus faibles, soit environ 39 % des pages. Cela est dû au fait que dans les domaines asiatiques, certaines langues régionales, dont l'hébreu, le thai ou le turc, sont dominantes dans leurs ccTLD respectifs.

En 2010 l'enquête fut élargie aux domaines de la Caraïbe et établit que la langue la plus utilisée, avec un ratio de 55 %, était l'espagnol. L'anglais venait en second avec environ 33 % des pages.

Connaissance et multilinguisme

Notre étude dégage un résultat significatif concernant la présence des connaissances de secteurs professionnels et pratiques sur le web selon les langues. Nous avons analysé la présence de 100 mots des champs scientifiques et techniques dans l'encyclopédie en ligne Wikipédia. Le terme *nitrogen* (azote) fut trouvé dans 171 langues différentes. Si l'ensemble des mots fut bien présent en anglais, parmi les autres langues européennes, le nombre moyen de mots présents était de 30. Dans les langues asiatiques, 18 mots en moyenne étaient référencés, tandis que dans les langues africaines, cette moyenne était seulement de 7.

Tableau 1. Référencement de 100 termes scientifiques et ingénieriques sur Wikipédia, par langue

Nombre de mots présents N	Langues européennes*	Langues asiatiques	Langues africaines	Autres langues
0-9	0	2	0	0
10-19	31	30	6	8
20-29	10	8	1	0
30-39	4	4	1	1
40-49	3	2	0	0
50-59	11	1	0	0
60-69	7	4	0	0
70-79	5	4	0	0
80-89	6	1	0	0
90-99	2	0	0	0
100	1	0	0	0
Nbre moyen des mots présents	30	18	7	15
Nbre de langues prises en compte	80	56	8	17

*Les langues européennes incluent l'anglais

On affirme souvent que la révolution internet va rendre universel l'accès à la société de la connaissance ; en réalité, nous découvrons d'importants écarts entre les langues pour ce qui concerne l'accès à des connaissances professionnelles. Une description plus détaillée des résultats de cette étude peut être consultée dans [HTUN ET AL. 2010].

Les initiatives de l'Unesco

Puisque la langue est un pourvoyeur essentiel de la culture, la diversité linguistique est un sujet incontournable dans des discussions plus larges concernant la diversité culturelle. Depuis sa création L'Unesco s'est engagée dans une démarche de préservation de la diversité culturelle reposant sur l'idée que « *le dialogue interculturel et le respect de la diversité culturelle et de la tolérance sont indispensables pour ériger une paix durable* » [UNESCO 2003]. L'Unesco n'a depuis cessé de publier des Déclarations et Recommandations relatives à la diversité culturelle.

En octobre 2003, les États membres de l'Unesco adoptèrent la *Recommandation sur le cyberspace*, affirmant l'importance de la diversité culturelle et soulignant la responsabilité de l'Unesco dans son maintien. L'Unesco a désormais une responsabilité dans le développement des contenus et systèmes multilingues, ainsi que des contenus du domaine public, et dans la facilitation de l'accès aux réseaux et services. Dans le cadre de cette Recommandation, diverses activités dont le SMSI (Sommet Mondial sur la Société de l'Information), le FGI (Forum sur la Gouvernance de l'Internet), et la Journée internationale de la Langue maternelle (JILM) ont été planifiées et mise en œuvre.

L'OBSERVATOIRE DES LANGUES

Le *Language Observatory Project* fut créé en 2003 dans la foulée de la publication par l'Unesco de sa *Recommandation sur le Cyberespace*. L'objectif premier du projet est l'observation et la fourniture de données sur l'état réel de l'utilisation des langues sur le Web, et ce afin de mesurer la diversité linguistique sur la Toile.

Le fonctionnement de l'observatoire

L'observatoire des langues est conçu pour mesurer l'utilisation de chaque langue sur le World Wide Web. Les mesures sont effectuées au travers du décompte de pages écrites dans chaque langue sur le Web.

Le projet repose sur deux composantes essentielles. La première est un instrument de collecte des données du web utilisant des robots d'indexation et qui, doublé d'un logiciel de routage parallèle hautement performant développé à l'université de Milan, [MIKAMI ET AL. 2005] peut collecter des millions de pages chaque jour.

Le second élément est un instrument d'identification des langues. L'observatoire a développé un logiciel permettant d'identifier de manière précise et à grande échelle les langues, les écritures et les propriétés d'encodage utilisés sur les pages web. La première version de l'algorithme d'identification LIM (Module d'Identification Linguistique) fut développée par Suzuki *et al.* en 2002 [SUZUKI ET AL. 2002] et implémenté par Chubachi *et al.* en 2004. Il fut plus tard amélioré par Chew en 2008 pour une seconde version qui correspond à celle actuellement utilisée, G2LI.

G2LI est capable d'identifier 184 langues de la liste des codes de langues ISO (ISO 639-1) avec une précision de 94 % au regard des dernières performances évaluées. En plus de couvrir un large ensemble de langues, il peut identifier différents types de codages de caractères non-standard³, qui sont encore souvent utilisés massivement par les communautés n'utilisant pas l'alphabet latin, comme indiqué précédemment. La seconde version de l'algorithme utilise des procédures améliorées de pré-compilation et peut gérer correctement le codage d'entités HTML⁴, utilisé abondamment également dans les écritures non-latines. Du fait de ces propriétés spécifiques, les auteurs estiment que le G2LI est l'instrument d'identification des langues le plus adapté à la mesure de présence de chaque langue sur le Web.

3 Il s'agit de système souvent propriétaires et non standardisés d'encodage des caractères qui datent d'avant la généralisation d'Unicode.

4 Les entités HTML sont un moyen d'écrire avec les seules lettres du code ASCII un ensemble plus vaste de caractères (par exemple, l'entité `&a1pha`; représente le caractère grec α).

Une composante invisible : La Déclaration Universelle des Droits de l’Homme

Au cœur de l’instrument d’identification des langues se cache un ensemble de textes ayant servi pour permettre au logiciel d’apprendre à reconnaître les langues. Les aspects techniques sont donnés dans [SUZUKI ET AL. 2002], mais il est important de préciser que la richesse et la qualité des textes de référence sont un point crucial de la tâche d’identification linguistique. Pour cela, nous avons utilisé un ensemble de textes traduits à partir de la Déclaration Universelle des Droits de l’Homme (DUDH), fournis par la Commission des Droits de l’Homme des Nations Unies, en raison de leur disponibilité en un très grand nombre de langues.

Autre précision d’importance, les textes traduits de la Déclaration Universelle des Droits de l’Homme n’ont pas tous été transmis encodés ; certaines versions n’existent que sous la forme de fichiers images. Ces images peuvent être lues par les êtres humains, mais pas directement par les ordinateurs, nécessitant leur transformation en données textuelles.

Tableau 2. Nombre de textes de la DUDH disponible sur le site de la Commission des Droits de l’Homme des Nations Unies par format.

	latin	cyrillique	autres alphabets	abjad	abugida	hanzi	tous les autres	Total
encodés	253	10	1	1	0	3	0	268
PDF	2	4	2	3	10	0	4	25
GIF	1	3	0	9	15	0	1	29
Total	256	17	3	13	29	3	5	322

NOTE : **Autres alphabets** : grec, arménien et géorgien ; **abjad** : arabe et hébreu ; **abugida** : amharic et tous les systèmes d’écriture brahmique utilisés dans le Sud et le Sud-Est de l’Asie ; **hanzi** : chinois, japonais et coréen ; **Tous les autres** : assyrien, syllabaires autochtones canadiens, ojibwa, cree, mongol et yi.

Le tableau 2 donne à voir combien de textes ont été donnés en fichiers images (322 langues étaient disponibles à la date de la première enquête, début 2004). Plus de deux cents langues utilisent l’alphabet latin, avec ou sans signes diacritiques, et seules trois d’entre elles furent transmises sous un format de fichier PDF ou GIF. À l’opposé, parmi les traductions

en langues utilisant l'écriture abugida⁵, aucune ne fut donnée sous la forme d'un texte codé.

Ce fait pourrait à lui seul illustrer l'existence d'une fracture numérique linguistique, ou dans ce cas précis, d'une « fracture numérique scripturale ». Après avoir constaté ce problème, l'un des auteurs rédigea un essai pour le journal indien *Vishbha Bharat*:

« J'ai récemment parcouru le site internet de la Commission des Droits de l'Homme des Nations Unies⁶ qui propose plus de trois cents versions linguistiques différentes – de l'abkhaz au zulu – de la Déclaration Universelle des Droits de l'Homme. Le site proclame que ce texte est celui qui a été le plus traduit au monde, et qui l'a obtenu son entrée dans le livre Guinness des Records pour une telle performance. Aussi la Déclaration Universelle est considérée comme « le texte le plus universel » au monde.

Vérifiez par vous-même! Et vous trouverez bien une version dans les dix-huit langues officielles de l'Inde, à deux exceptions près – le konkani et le manipuri. Mais la frustration serait grande quand vous constaterez que l'ensemble de ces versions en langues indiennes sont diffusées sous la forme de fichiers.gif, et non pas sous la forme de textes encodés. Un grand nombre d'utilisateurs d'écritures non latines dans le même cas doivent ressentir le même genre de tristesse après leur consultation du site. [MIKAMI 2002]

Depuis ce témoignage, de nombreux collaborateurs nous ont apporté leur aide pour créer des versions encodées de ces fichiers images⁷. Pour certaines langues nous cherchons toujours des contributeurs et avons

5 Les écritures abugida, ou alphasyllabaires, sont des écritures syllabiques, dont la plupart sont issues des écritures brahmiques indiennes et utilisées dans les régions sud et sud-est asiatiques. Une écriture abugida majeure est l'amharique.

6 <http://www.unhchr.ch/udhr>

7 sinhala, vietnamien, bahasa melayu, lao, persan (farsi), mongol, tamil, ouïgour, népalais, malayalam, hindi, magahi, marathi, sanskrit, bengali, saraiki, punjabi, gujarati, kannada, myanmar, vietnamien en TCVN5712, VIQR, VPS, assaais, azéri, dari, kirghiz, marwari, sindhi, tajiki, tamang, telugu, turkmène, urdu, ouzbek. Ces textes furent encodés en UTF-8, à l'exception des langues indiquées... Même si UTF-8 reste encore limité pour certaines langues et si de nombreuses langues utilisent également d'autres systèmes propriétaires de scripts.

Pour des compléments et la liste des contributeurs, voir notre site: <http://gii2.nagaokaut.ac.jp/gii/lopdary.php?itemid=480>

du renoncer pour l’instant à leur inclusion dans les textes destinés à l’apprentissage du système.

Approximativement à l’époque où nous initions le *Language Observatory Project*, Eric Miller a démarré le projet UDHR-in-Unicode. L’objectif était de donner à voir l’utilisation de l’Unicode par une large palette de langues, en utilisant la *Déclaration Universelle des Droits de l’Homme* comme texte représentatif. Le projet UDHR-in-Unicode est actuellement hébergé sur le site web du Consortium Unicode et les textes sont utilisés pour la recherche sur le traitement du langage naturel⁸.

Sponsors et collaborateurs

Le *Language Observatory project* fut initié par les auteurs de cet article en 2003 et bénéficia d’aides financières de 2003 à 2007 en provenance de la Japan Science and Technology Agency (JST) via son programme RISTEX. L’inauguration eut lieu à la Nagaoka University of Technology en février 2004 en présence de Paul Hector, alors directeur de la section Communication et Information de l’Unesco.

Le projet s’est développé en collaboration et interaction avec de nombreux partenaires partout dans le monde, notamment l’Académie Africaine des Langues (ACALAN) lors du SMSI de Tunis en Novembre 2005 durant une session de travail sur les langues africaines. Parmi les participants, se trouvaient le président de l’ACALAN, Adama Samassekou, Daniel Pimienta de Funredes, et Daniel Prado de l’Union Latine.

D’un commun accord, nous avons décidé d’organiser conjointement un projet d’enquête sur les langues africaines en ligne. La cible initiale du projet était à l’origine les domaines ccTLD africains. En 2006, nous avons tenu un atelier à Bamako, au Mali, avec la coopération de l’ACALAN et le soutien du JST. Un grand nombre de chercheurs africains concernés par la diversité linguistique et la fracture numérique ont participé à ce séminaire.

Après cet atelier prometteur à Bamako, nous avons organisé diverses réunions pour promouvoir notre projet et sensibiliser à la fracture numérique linguistique, ainsi que des ateliers au siège de l’Unesco lors des Journées internationales de la Langue maternelle de 2007 et 2008.

⁸ NLTK (*Natural Language Tool Kit*), de Steven Bird et al. en est un exemple.

Le rapport de l'Observatoire, publié en 2008 [NANDASARA ET AL. 2008], fut le premier article posant la question de la distribution des langues sur le web asiatique. Il mit en évidence une fracture numérique linguistique significative, avec l'anglais utilisé dans plus de 60 % des pages web en Asie du Sud et du Sud-Est. Dans les régions Ouest de l'Asie, la domination anglaise était moins outrancière, et dans certains pays, l'arabe était la langue la plus courante. En Asie centrale, le russe restait la langue dominante, sauf au Turkménistan où l'anglais était utilisé pour 90 % des pages. Une minorité de langues locales, dont le turc, l'hébreu, le thaï, l'indonésien, le vietnamien et le mongol étaient les langues les plus utilisées dans leurs propres domaines nationaux. Cette étude marqua une étape importante pour la compréhension de la disparité linguistique en ligne, et nous offrit une base solide pour les travaux suivants.

LA DIVERSITÉ LINGUISTIQUE SUR LE WEB

Dans cette section, quelques résultats des enquêtes menées par l'observatoire des langues seront présentés.

L'indice de diversité de Lieberson

L'indice de diversité de Lieberson (LDI) [LIEBERSON 1981] est un indice de diversité linguistique largement utilisé et qui est défini par la formule suivante, où P_i représente la part respective des (*i* comme autant de langues) locuteurs d'une même langue dans une communauté :

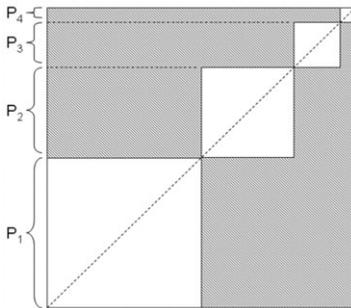
$$\text{LDI} = 1 - \sum P_i^2$$

Si les personnes de la communauté parlent toutes la même langue alors $P_i = 1$ alors que pour les locuteurs d'une autre langue, $P_i = 0$. Aussi, le LDI d'une communauté complètement monolingue est zéro. Si 4 langues sont parlées par un nombre égal de personnes, alors $P_1 = P_2 = P_3 = P_4 = 0.25$ et le LDI de cette communauté multilingue peut être calculé comme suit : $\text{LDI} = 1 - (0.25)^2 \times 4 = 0.75$. Ainsi, plus le LDI est élevé, plus la diversité linguistique est importante, et inversement.

Lieberson prit également en compte le fait que des locuteurs bilingues ou multilingues compliqueraient la formule. Mais l'idée maîtresse du LDI peut être comprise au travers de l'illustration en figure 1. Un carré

de largeur P_i représente la probabilité qu'a tout locuteur d'une langue de rencontrer un locuteur de la même langue. Et la somme des carrés de P_i représente la probabilité combinée que n'importe lequel des locuteurs rencontre un locuteur de la même langue dans la communauté. Enfin, la somme des carrés de P_i est retranchée de 1, indiquant la probabilité que n'importe lequel des locuteurs rencontre des locuteurs d'autres langues dans la société. Les aires pleines du carré en figure 1 correspondent à cette probabilité.

Figure 1. une interprétation graphique de LDI

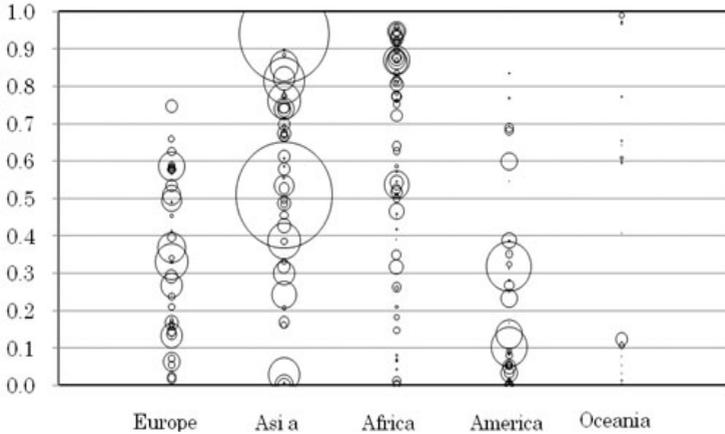


Le répertoire *Ethnologue* fournit une liste complète des Indices de Diversité de Liberson par pays ou région, ainsi que la taille de la population et le nombre de langues indigènes ou d'immigrants. À partir de ces données⁹, les auteurs construisirent la Figure 2 pour montrer comment le LDI change d'un pays ou d'un continent à l'autre. Chaque cercle correspond à un pays. La surface du cercle rend compte de la population du pays, et l'axe vertical représente le LDI de chaque pays. Les deux grands cercles sur l'axe de l'Asie correspondent à l'Inde (LDI = 0.94) et à la Chine (LDI = 0.51).

Comme le montre ce graphique, les pays du continent africain affichent la plus grande diversité linguistique parmi l'ensemble des continents, suivis par l'Asie, l'Europe, les Amériques (Sud et Nord réunis) et l'Océanie.

⁹ Diffusées dans la version en ligne, soit l'équivalent de la 16^e édition de *Ethnologue*.

Figure 2. L'indice de diversité de Lieberson dans chaque pays et par continent



(À partir de données provenant de *Ethnologue*)

L'Indice de Diversité de Lieberson le plus élevé se trouve dans la République Centrafricaine (LDI = 0.96); neuf autres pays ont un LDI supérieur à 0.90 (la République Démocratique du Congo, la Tanzanie, le Cameroun, le Tchad, le Mozambique, l'Ouganda, le Bénin, la Côte d'Ivoire et le Libéria). Treize pays dépassent les 0.8 (Le Togo, la Zambie, Le Kenya, l'Afrique du Sud, le Mali, la Guinée-Bissau, le Nigéria, l'Éthiopie, le Congo, la Sierra Leone, l'Angola, la Namibie et le Ghana), et 17 autres ont un LDI dépassant 0.5. Les pays au LDI le plus bas en Afrique sont le Rwanda et le Burundi, avec 0.004.

En Asie, la diversité la plus élevée est observable en Papouasie Nouvelle Guinée (0.99). Ce pays est connu pour ses ressources linguistiques abondantes, et son LDI est le plus haut de la planète; il est suivi sur le continent asiatique par l'Inde, le Timor Oriental, le Bhoutan, les Philippines, l'Iran et l'Indonésie. Ces sept pays ont un LDI supérieur à 0.8, tandis que 22 autres pays asiatiques passent la barre des 0.5. À l'opposé sur l'échelle de la diversité selon l'indice de Lieberson, la Corée (0.003), les Maldives (0.01) et le Japon (0.03) apparaissent comme des sociétés fortement monolingues.

En Europe, les LDI les plus élevés sont à chercher en Belgique (0.75), en Bosnie (0.66) et en Serbie (0.63). Viennent ensuite la Moldavie (0.59), l'Italie (0.59), la Lettonie (0.58), la Géorgie (0.58), la Macédoine (0.58), la

Suisse (0.58), l'Albanie (0.57), Andorre (0.57), l'Autriche (0.54), Monaco (0.52), et l'Espagne (0.51). Ces 15 pays ont un LDI supérieur à 0.5. Des pays caractérisés une langue première dominante comme l'Allemagne (0,37), la Russie (0,33), les Pays-bas (0.29), et la France (0.27), ont généralement de faibles LDI. Avec la Hongrie en queue de peloton en Europe (0.02).

Sur le Continent américain, seuls trois pays ont un LDI supérieur à 0.5 : Belize (0.77), Trinidad et Tobago (0.70), et le Canada (0.60). Les pays dont la langue dominante est l'espagnol ont majoritairement un faible LDI.

En Océanie, la séparation des populations sur de petites îles isolées est le terreau d'un développement monolingue par île. Des pays composés de plusieurs îles tendent à avoir un LDI plus élevé. Le LDI du Vanuatu est de 0.97 et le plus haut parmi les pays de l'Océanie; plus de 100 langues y sont parlées. D'autres pays composés d'archipels présentent un LDI élevé: les Îles Solomon (0.97), la Nouvelle-Calédonie (0.83), la Micronésie (0.77), les Fidji (0.61), et Nauru (0.60).

Ratio linguistique au niveau local

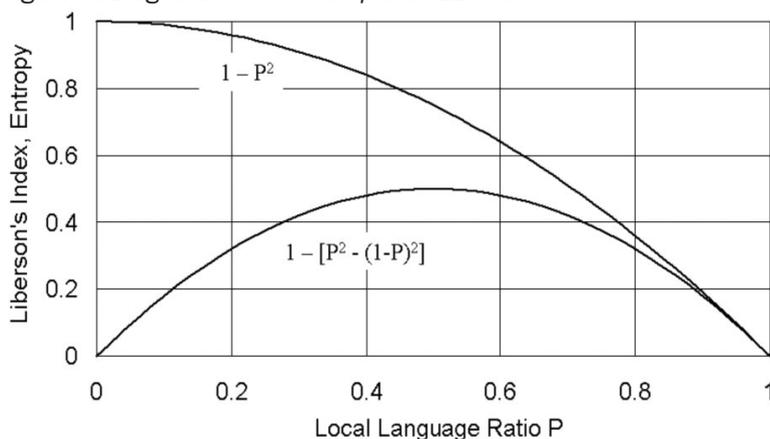
Dans le chapitre précédent, nous avons dressé un panorama global de la diversité linguistique dans le monde à partir de données fournies par *Ethnologue*, des données qui reflètent la situation *dans le monde réel*. Nous souhaitons désormais aborder la partie principale de cet article, soit la diversité linguistique dans le *cybermonde*.

Depuis sa création, le *Language Observatory Project* a concentré son attention sur deux continents, l'Asie et l'Afrique. Comme mentionné plus haut, les premiers résultats observés ont été diffusés lors d'un atelier organisé au siège social de l'Unesco en février 2005; ils ont fait l'objet d'un article publié en 2008. L'observatoire a récemment conduit d'autres séries d'enquêtes en Asie, en Afrique et dans la région des Caraïbes à partir de données datant de 2009. Les chapitres suivants présenteront un aperçu de cette très récente étude.

Nous proposons ci-dessous un graphique à deux axes, que nous avons intitulé de manière expérimentale le LL-chart; le Ratio Linguistique Local est figuré sur l'axe horizontal et le LDI sur l'axe vertical. Le but de ce graphique est de résoudre un problème que nous avons rencontré en construisant un graphique sur les LDI à partir de données issues du

cyberespace. Il n'est en effet pas rare de constater que les langues utilisées sur le web sont complètement différentes de celles parlées dans le monde réel. Dans un grand nombre de cas, les langues parlées sont des langues locales alors que les langues du web sont principalement des langues « globales », comme l'anglais, le français ou le russe. Et dans ces cas-là, les LDI élevés concernant les langues dans le cyberespace et ceux dans le monde réel ne sont pas considérés comme équivalents. Nous devons prendre en compte certaines mesures concernant la présence de langues locales, comme présentée dans la figure 3.

Figure 3. Diagramme schématique du LL-chart



Il est à noter que tous les pays disposant d'un Ratio Linguistique Local P sont représentés dans la zone entre les deux courbes $1 - [P^2 + (1 - P)^2]$, soit l'indice de Lieberston dans le cas où l'on a deux langues, et $1 - P^2$, qui donne la valeur maximale de l'Indice de Diversité de Lieberston¹⁰. Lorsque P est supérieur à 0.5, le LDI diminue et le point de la courbe se déplace vers le coin inférieur droit. Lorsque P est petit, deux possibilités se présentent : soit l'absence d'une langue locale est comblée par une langue étrangère dominante, entraînant la diminution du LDI et un positionnement en bas à gauche du graphique ; soit elle est comblée par plusieurs autres

10 Deux courbes correspondent aux limites inférieure et supérieure. La courbe supérieure indique le LDI d'une communauté à deux langues. Comme l'ajout d'un locuteur d'une troisième langue augmente la probabilité de rencontrer des locuteurs de langue différente, cette valeur est le LDI minimum de communautés ayant plus de deux langues. La courbe inférieure indique le LDI d'un cas très spécifique, où chaque membre, en plus de la langue locale, parle une langue autre, soit le LDI maximum.

langues, et dans ce cas le LDI augmente et le point se déplace en haut à gauche dans le schéma.

Comparaison par région : Asie, Afrique, Europe et les Caraïbes

À partir de données collectées en novembre 2009, le LDI et le Ratio Linguistique Local ont été calculés pour tous les domaines nationaux en Asie et en Afrique. Comme nous n'avons pas de données pour les pays européens, nous avons utilisé le décompte de pages par langue publié par Google. Les Figures 4, 5 et 6 présentent le graphique combinant Ratio Linguistique Local et LDI pour ces trois régions.

Les LDI asiatiques sont visibles sur la figure 4. La Chine, le Japon, la Corée et certains pays arabophones (Iraq, Arabie Saoudite, et Jordanie) se situent dans le coin inférieur droit, tandis que le Vietnam, la Thaïlande et l'Indonésie, Israël, la Turquie, la Géorgie et la Mongolie affichent une présence relativement élevée des langues locales.

Autre élément remarquable, le contexte des pays d'Asie Centrale. Leurs espaces web sont composés en langues locales, avec toutefois des composantes clés en anglais ou en russe, selon le pays. Le Kazakhstan, le Kirghizistan, le Tadjikistan, et l'Ouzbékistan mettent l'accent sur le russe, tandis que seul le Turkménistan a une préférence pour l'anglais.

Toutefois, les contenus web du sous-continent Indien montrent une présence quasi négligeable de langues locales. Plus de 70 % de ces contenus web sont écrits en anglais.

Le cas du Laos est particulier et mérite qu'on s'y arrête ici. Selon *Ethnologue*, le LDI du pays est seulement de 0.674. Pourquoi dès lors a-t-il un LDI aussi élevé sur le Web ? La raison principale réside dans le fait que le domaine `.la` est activement promu auprès des étrangers, dont certains se connectent même de Los Angeles. Comme le domaine est essentiellement vendu à des industries et personnes étrangères, seulement 8 % des pages en `.la` sont en langue lao.

Les LDI des domaines africains sont représentés en figure 5. La présence des langues locales dans les domaines africains est beaucoup plus rare que dans les domaines asiatiques. Pour les pays arabophones, la langue locale emporte la majorité seulement au Soudan et en Libye ; L'Égypte,

La Mauritanie, la Tunisie et la Tanzanie, ainsi que le reste de l’Afrique affichent une présence faible de leurs langues locales sur le Web. Cependant, plusieurs pays n’en ont pas moins des LDI élevés sur le Web.

Les LDI des domaines européens et de certains domaines anglophones sont en figure 6. La présence linguistique locale est au dessus de 50 % à l’exception de la Slovénie et du Danemark, dont les espaces web sont dominés par l’anglais, entraînant un LDI plus faible. De l’autre coté du spectre, le Royaume-Uni, affiche un LDI très bas caractéristique d’autres pays anglophones (USA, Australie et Nouvelle-Zélande).

Table 3. La composition linguistique des domaines web asiatiques et africains

Domaines africains			Domaines asiatiques		
Langue	nb de pages	%	Langue	nb de pages	%
anglais	30 327 396	78,40%	chinois	7 832 521	20,46%
français	2 737 455	7,08%	japonais	5 287 655	13,82%
afrikaans	660 510	1,71%	anglais	4 867 355	12,72%
arabe	592 746	1,53%	russe	1 611 339	4,21%
chinois	391 745	1,01%	coréen	1 100 232	2,87%
portugais	348 131	0,90%	vietnamien	710 048	1,86%
russe	307 178	0,79%	thaï	544 561	1,42%
espagnol	276 126	0,71%	indonésien	308 894	0,81%
japonais	158 992	0,41%	hébreu	89 076	0,23%
autres	879 605	2,27%	autres	14 055 334	36,72%
non identifiée	2 005 311	5,18%	non identifiée	1 867 355	4,88%
Total	38 685 195	100,00%	Total	38 274 370	100,00%

NOTE : Les données Web ont été récupérées des domaines nationaux en Asie et en Afrique en Novembre 2009. Pour une liste complète des domaines, voir les annexes.

Figure 4. Ratio Linguistique Local et LDI de la composition linguistique des domaines asiatiques.

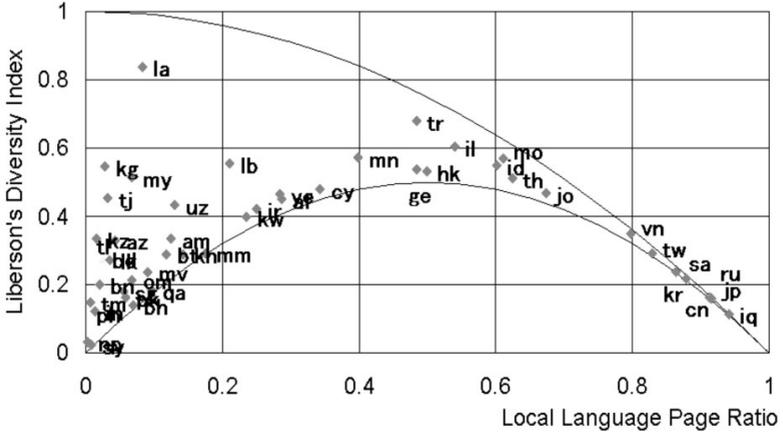


Figure 5. Ratio Linguistique Local et LDI de la composition linguistique des domaines africains.

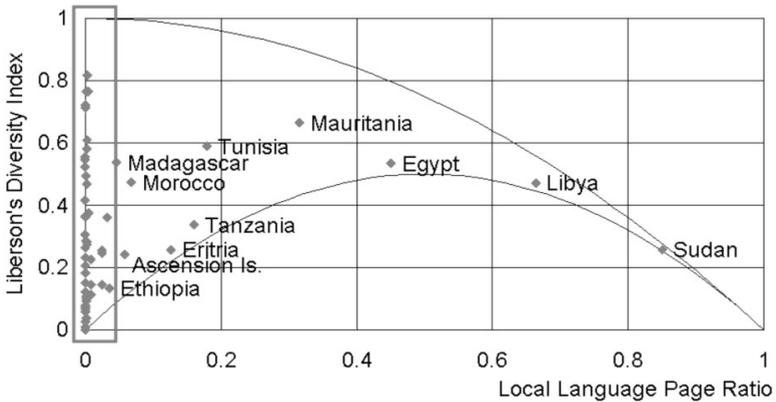
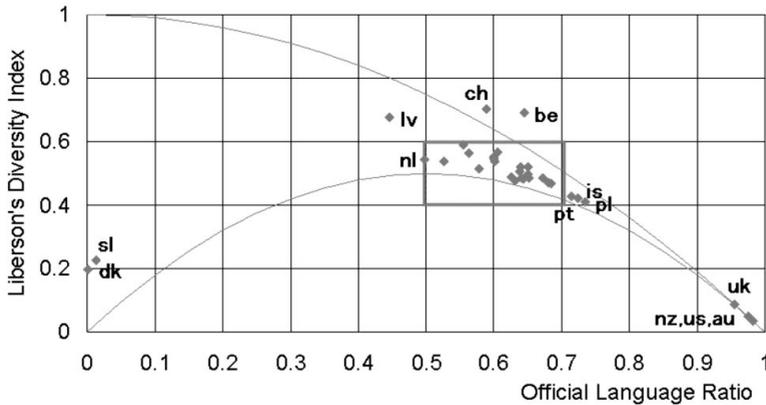


Figure 6. Ratio Linguistique Local et LDI de la composition linguistique sur le Web européen pour les pays anglophones sélectionnés



Challenges et orientations

Le défi le plus important posé par ces mesures provient de la taille d'un web qui ne cesse de grandir. Personne ne sait exactement combien de pages existent dans le World Wide Web. En 1997, le nombre avait été estimé à seulement 320 millions; dès 2002, il avait atteint 8 milliards [MILLER 2007]. En 2008, Google annonça mille milliard d'URL sur le Web, mais n'a depuis fourni aucune autre donnée. Et aucun autre moteur de recherche n'en procure, ce qui nous mène à penser qu'il est actuellement impossible de comptabiliser l'ensemble des pages web existantes.

Une autre stratégie est nécessaire pour créer une méthode d'échantillonnage de pages qui refléterait le Web dans son ensemble. Nous sommes en cours de développement d'une méthode qui nous semble prometteuse à l'aide de l'ANOVA (Analyse de la variance.)

Un autre avantage de l'échantillonnage est de pouvoir étendre la cible de recherche à d'autres ccTLD qui n'avaient pas été pris en compte en raison de leur taille gigantesque.

Nous avons transmis des rapports annuels faisant état de ces statistiques d'utilisation des langues sur le Web lors d'événements organisés par

l'Unesco ou le FGI, et nous continuerons à les faire parvenir dans le futur, en y apportant les améliorations suivantes :

- Élargissement de la cible
- Élargissement du nombre de langues identifiables
- Diversification des méthodes d'analyse

La première amélioration a été évoquée plus haut. Les aires de recherche que nous ciblons incluent pour l'instant l'Asie, l'Afrique, les Caraïbes et l'Europe. Un grand nombre de ccTLD sont encore manquants dans notre recherche en raison de trop faibles capacités de stockage de notre système.

La seconde amélioration nous aidera à dessiner une image encore plus précise de l'utilisation des langues sur le Web. Notre moteur d'identification peut reconnaître plus de 300 langues, mais selon une estimation d'*Ethnologies*, plus de 7 000 langues existent sur Terre. Comme beaucoup d'entre elles n'ont pas une forme écrite et sont seulement parlées, comme indiqué dans le tableau 3, notre outil d'identification n'a pas pu identifier environ 5 % des pages collectées. Nous en avons donc conclu que nous négligions un grand nombre de langues. Comme nous l'avons indiqué dans la première partie de cet article, il nous faut collecter des codages locaux pour trouver les solutions aux problèmes d'identification des codages de caractères.

Un prototype de la troisième version a été présenté en partie 3. La donnée la plus basique que nous pouvons fournir est une liste du nombre de pages dans chaque langue pour chaque ccTLD. Mais ces informations ne vont pas nous dire grand-chose de l'utilisation réelle de ces pages, et donc de la place des langues sur la Toile. Nous devons affiner notre interprétation pour permettre une réflexion plus profonde sur l'utilisation des langues.

Forts de ces améliorations, nous espérons augmenter l'utilité des statistiques comme données fondamentales de la détermination de l'usage et de la diversité linguistique sur la Toile.

BIBLIOGRAPHIE

[PRIOLKAR 1958] A. K. Priolkar. 1958. *The Printing Press in India*. Bombay : Marathi Samsodhana Mandala.

[MIKAMI 2002] Yoshiki Mikami. 2002. Global digital-divide among scripts. *VishwaBharat*. October 2002 Issue, p.1.

- [PIMIENTA ET AL. 2010] Daniel Pimienta, Daniel Prado and Álvaro Blanco. 2010. *Twelve Years of Measuring Linguistic Diversity in the Internet*. Paris: Unesco.
- [HTUN ET AL. 2010] Ohnmar Htun, Shigeaki Kodama and Yoshiki Mikami. 2010. Analysis of Terminology Terms in Multilingual Terminology Dictionary. *Proceedings of the 8th International Conference on Computer Applications* 2010, pp. 122-128.
- [UNESCO 2003] *Recommendation Concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace*. Unesco: Paris.
- [ISC 2010] Internet Software Consortium. 2010. *Internet Domain Host Count* <http://www.isc.org/solutions/survey>
- [UNESCO 2001] *Universal Declaration on Cultural Diversity*. Unesco: Paris.
- [MIKAMI ET AL. 2005] Yoshiki Mikami, Zavorsky Pavol, Mohd Zaidi abd Rozan, Izumi Suzuki, Masayuki Takahashi, Tomohide Maki, Irwan Nizam, Massimo Santini, Paolo Boldi, and Sebastiano Vigna. The Language Observatory Project. 2005. *Proceedings of the 14th International World Wide Web Conference*, p. 990.
- [SUZUKI ET AL. 2002] Izumi Suzuki, Yoshiki Mikami, Ario Ohsato. 2002. A Language and Character Set Determination Method Based on N-gram Statistics. *ACM Transactions on Asian Language Information Processing*, Vol. 1, No.3. pp. 270-279.
- [NANDASARA ET AL. 2008] S. T. Nandasara, Shigeaki Kodama, Chew Yew Choong, Rizza Caminero, Ahmed Tarcan, Hammam Riza, Robin Lee Nagano and Yoshiki Mikami. 2008. An Analysis of Asian Language Web Pages. *The International Journal on Advances in ICT for Emerging Regions (ICTer)*, Vol.1 No.1. pp. 12-23.
- [LIEBERSON 1981] Stanley Lieberman and Anwar S. Dil. 1981. *Language Diversity and Language Contact: essays*. California: Stanford University Press.
- [MILLER 2007] Miller, Colleen. 2007. Web Sites: Number of Pages. *NEC Research*. IDC. 6 June 2007.

LES TECHNOLOGIES DE LA LANGUE EN SOUTIEN AU MULTILINGUISME

Les enjeux du multilinguisme sont nombreux, et les besoins importants, tant en Europe qu'à l'échelle internationale. Les technologies de la langue peuvent aider à y répondre, mais nécessitent de développer les infrastructures et de produire les ressources nécessaires à la conduite des recherches pour les différentes langues. Certains programmes soutiennent ce domaine, mais souffrent d'un manque d'ampleur, de continuité et de cohésion. Cet effort mériterait d'être coordonné entre les nations et les organismes internationaux, afin de faciliter le multilinguisme, en Europe comme au plan mondial.

Article original en français.



JOSEPH MARIANI est actuellement directeur de l'Institut franco-allemand des Technologies Multilingues et Multimédias de l'Information (IMMI). Il a été directeur du LIMSI-CNRS et responsable de son Département Communication Homme-Machine, puis directeur du Département des Technologies de l'Information et de la Communication au Ministère de la Recherche.

JOSEPH MARIANI

LES
TECHNOLOGIES
DE LA LANGUE
EN SOUTIEN AU
MULTILINGUISME

Depuis la punition divine de Babel, l'humanité doit vivre avec la richesse que procurent une multitude de langues et de cultures. La difficulté et le coût de partager l'information et de communiquer, malgré la barrière des langues tout en préservant ces langues, pourraient bénéficier d'un soutien majeur venant des systèmes de traitement automatique des langues, ou technologies de la langue, qui font l'objet d'un effort de recherche important, mais encore insuffisant et insuffisamment coordonné.

LES ENJEUX DU MULTILINGUISME

Les enjeux du multilinguisme comportent deux volets :

Tout d'abord, veiller à préserver les cultures et les langues, c'est-à-dire permettre aux citoyens, de s'exprimer dans leur langue maternelle. Cette question prend un relief particulier dans le cadre de la construction de l'Europe, compte tenu de la forte diversité linguistique dans un même ensemble politique. Ainsi, 75 % des citoyens Allemands interrogés préfèrent trouver des sites web dans leur propre langue que dans une langue étrangère. On peut également noter qu'actuellement seulement 30 % du Web est en anglais, pourcentage en forte diminution puisqu'il était de 50 % en 2000¹. Pour moitié, les citoyens européens ne parlent qu'une langue et quand ils en parlent une deuxième, ce n'est pas forcément l'anglais. À peine 3 % des Japonais parlent une langue étrangère. En Inde, moins de 5 % des citoyens parlent anglais. Préserver les langues et, à travers elles, les cultures répond de ce fait à une demande forte des citoyens.

1 Voir dans cet ouvrage : Michaël Oustinoff, *L'anglais ne sera pas la Lingua Franca de l'internet*.

Le deuxième défi est de permettre la communication entre les humains, en général comme dans le cadre de constructions démocratiques communes. Nous en vivons un exemple au niveau de l'Union Européenne, où l'on compte aujourd'hui 27 pays membres et 23 langues officielles, ce qui représente 506 paires de langues. Si l'on considère l'ensemble des langues européennes, on peut en compter plus de 60, ce qui représente près de 4000 paires de langues à traduire! La Commission Européenne emploie plus de 2500 traducteurs qui ont traduit en 2007 plus d'un million et demi de pages. Cela ne permet de couvrir qu'une partie des besoins. En couvrir la totalité nécessiterait 8500 traducteurs pour traiter annuellement 6,8 millions de pages. La prise en compte de la diversité linguistique représente 30 % du budget du Parlement Européen, soit de l'ordre de 300 millions d'euros par an, avec l'emploi de 500 traducteurs et interprètes. L'estimation du coût total du multilinguisme pour l'Union Européenne est d'un peu plus d'un milliard d'euros par an; mais rapporté au nombre de citoyens européens, cela ne représente jamais que 2,2 euros par citoyen et par an, ce qui finalement n'est pas rédhibitoire. Une telle situation existe également à l'intérieur de certaines nations, comme en Inde, mais aussi au plan international, avec de l'ordre de 6000 langues principales qui sont parlées, soit 36 millions de paires de langues à traduire... Et une simple statistique: à l'heure actuelle, sur YouTube, il se dépose vingt-quatre heures de vidéos nouvelles toutes les minutes, dans toutes les langues.

LES BESOINS LIÉS AU MULTILINGUISME

Au plan européen, les besoins liés au multilinguisme sont très nombreux: besoins pour la mise en place de la Bibliothèque Numérique Européenne (Europeana, qui en janvier 2011, comprenait 14,6 millions de documents dans 26 langues), pour laquelle il faut disposer d'outils multilingues et interlingues afin de permettre l'accès à tous; pour la réalisation projetée par l'Agence de Sécurité Européenne (ENISA) d'une plate-forme multilingue d'alerte et d'échange d'informations pour les États Membres; pour l'Office Européen des Brevets – le Protocole de Londres a réduit le nombre de langues officielles à trois (l'anglais, l'allemand et le français) pour des questions de coût, alors qu'avec plus d'outils automatisés, on pourrait traiter plus de langues; pour les réunions de la Commission Européenne,

du Parlement Européen ou de la Cour de Justice Européenne, où l'anglais tend de plus en plus à devenir seule langue de travail...

De tels besoins répondant à une véritable nécessité démocratique se retrouvent plus généralement au plan international. Si nous prenons l'exemple de la gouvernance de l'Internet, au sein du Forum de l'ONU sur la Gouvernance de l'Internet (FGI), seul l'anglais est admis comme langue de travail et un vif débat concerne la possibilité d'utiliser différentes graphies et différents accents dans les noms de domaine. La Bibliothèque Numérique Mondiale de l'Unesco comporte 1500 documents dans sept langues déposés lors de sa création en avril 2009². Le doublage et le sous-titrage d'œuvres audiovisuelles; la rédaction de notices techniques, dans l'aéronautique ou l'automobile, ou de modes d'emplois; le surtitrage d'œuvres du spectacle vivant; la traduction des textes, des vidéos, et des émissions de radio ou de télévision qu'on trouve de manière innombrable, dans toutes les langues; l'aide à l'interprétation simultanée dans les multiples réunions, conférences, ateliers, cours, qui ont lieu de par le monde; autant d'applications dans lesquelles les technologies de la langue peuvent offrir des opportunités. Pensons aussi aux besoins criants liés à la rédaction des articles scientifiques dans la langue maternelle, qui diminue fortement en raison d'une survalorisation par la bibliométrie de l'anglais au risque de voir disparaître les terminologies spécialisées dans les autres langues.

Ajoutons à ce panorama les nombreux besoins liés à l'accessibilité de l'information aux personnes handicapées, malvoyantes ou malentendantes, nécessitant la traduction de l'information d'un média dans un autre (de l'écrit à l'oral, de l'oral à l'écrit, de l'oral au geste (langue des signes)), et plus généralement à l'accessibilité de l'information aux personnes qui ne parlent pas la langue dans laquelle elle a été encodée, notamment les personnes migrantes³.

² <http://www.wdl.org/fr>

³ Voir dans cet ouvrage : Viola Krebs & Vicent Climent-Ferrando, *Langues, cyberspace et migrations*.

CONSTATS

L'ampleur de ces besoins montre bien qu'ils ne peuvent pas être couverts par les ressources humaines actuelles et même futures des métiers traitant de la langue.

La prise en compte du multilinguisme ne constitue pas une première priorité dans aucun secteur économique. Si on demande au patron d'une grande société quelle est sa priorité, aucun ne dira que c'est le multilinguisme. Mais si on fait le total des priorités dans chacun des secteurs où sa prise en compte est nécessaire, alors on arrive à une somme très importante. Cela nécessite donc, à notre avis, une réflexion et une action politiques pour amener à cette prise de conscience et apporter les réponses adéquates.

Si le multilinguisme apparaît comme une nécessité, son coût est cependant très important. C'est cet écart qui plaide pour le développement des technologies de la langue, et leur utilisation quand leurs performances sont à la hauteur des besoins des applications visées.

Il faut souligner qu'actuellement, les technologies de la langue n'ont pas encore atteint leur maturité pour toutes les langues, avec de forts déséquilibres selon les langues. Et qu'elles ne dispensent nullement d'une intervention humaine. Ainsi, la traduction automatisée n'a pas une qualité suffisante pour traduire des œuvres littéraires ou, en général, des textes qui nécessitent une traduction de qualité. Il faut l'affirmer clairement. Mais, par contre, elle peut aider le traducteur humain dans son activité et possède une qualité suffisante pour donner une traduction approximative, de pages web par exemple, répondant ainsi aux besoins du grand public. Les technologies de la langue participent pleinement à traiter la question du multilinguisme, ce qui justifie d'y apporter toute l'attention qu'elles méritent, notamment en matière de financement de programmes de recherche.

LES TECHNOLOGIES DE LA LANGUE

Les technologies de la langue sont dites *monolingues*, quand elles traitent une seule langue, *multilingues*, quand une même technologie en traite plusieurs, ou *interlingues*, quand elles permettent de passer d'une langue à une autre.

Nombre de technologies de la langue concernent le traitement de la langue écrite, qu'elles soient monolingues (analyse morphosyntaxique et syntaxique, compréhension, génération de textes, résumé automatique, extraction de terminologie, recherche d'information, systèmes de Réponse aux Questions...), ou interlingues (traduction automatique ou assistée par ordinateur, recherche d'information interlingue...).

Pour l'oral, on trouve également des technologies monolingues : reconnaissance et compréhension de la parole, transcription – reconnaissance et restitution sous forme écrite de ce qui est dit –, synthèse vocale, dialogue oral, reconnaissance de la personne qui parle – le locuteur – ; et interlingues (identification de la langue parlée, traduction vocale, interprétation simultanée...).

Enfin, il ne faut pas non plus oublier la communication gestuelle, en particulier pour le traitement des langues des signes (reconnaissance, synthèse et traduction)⁴.

Ces technologies peuvent être intermédias, permettant de passer d'un média à un autre, avec de nombreuses applications pour permettre l'accessibilité des personnes handicapées : synthèse vocale à partir du texte pour les déficients visuels, transcription automatique (sous-titrage ou surtitrage) ; aide à la lecture labiale, traitement des langues des signes..., pour les déficients auditifs ; commande vocale pour les déficients moteurs...

En sciences et technologies de la langue, les recherches recouvraient initialement deux domaines relevant de deux communautés scientifiques différentes :

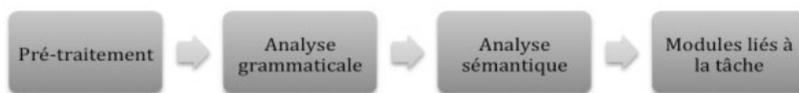
- Le traitement du langage écrit (traitement automatique de la langue [TAL], ou traitement du langage naturel), venant de la linguistique et de l'intelligence artificielle ;
- Le traitement du langage parlé (intitulé « communication parlée »), venant de l'acoustique, du traitement du signal, de la reconnaissance des formes.

Ces deux communautés se rejoignent progressivement, du fait d'une volonté politique et de l'utilisation de méthodes voisines fondées sur l'apprentissage automatique, utilisant des modèles statistiques.

⁴ Voir dans cet ouvrage : Annelies Braffort & Patrice Dalle, *Accessibilité du cyberspace : langues et handicap*

Les recherches dans ces deux grands domaines ont beaucoup progressé sur les niveaux dits inférieurs du traitement : pour l'écrit, segmentation, analyse lexicale, analyse morpho-syntaxique et syntaxique ; et pour l'oral, reconnaissance vocale, synthèse vocale à partir du texte, reconnaissance du locuteur. Nombre d'applications qui en découlent sont aujourd'hui d'usage courant : correcteurs orthographiques et syntaxiques, moteurs de recherche monolingue et interlingue, traduction automatique en ligne pour l'écrit, GPS parlant, systèmes de dictée vocale, transcription et indexation automatique de contenus audiovisuels pour l'oral... On constate ainsi que nombre de ces applications sont liées aux relations entre parole et écrit (transcription de parole, synthèse à partir du texte). Les systèmes de dialogue oral, incluant reconnaissance et synthèse vocales, se développent également, mais dans le cadre d'applications très spécifiques : commande vocale sur les téléphones mobiles, menus de centres d'appels, informations touristiques ou relatives aux transports en commun...

Architecture de base d'un système de traitement de la langue naturelle⁵



Le domaine de la recherche en traduction automatique illustre particulièrement bien la rencontre de ces deux communautés. Ce domaine était traditionnellement traité par les chercheurs en TAL, utilisant des approches par règles et des connaissances linguistiques (dictionnaires bilingues, grammaires...). Les chercheurs travaillant dans le domaine de la communication parlée ont pour leur part expérimenté en traduction automatique les méthodes d'apprentissage automatique qu'ils ont utilisées avec succès en reconnaissance vocale, pour mettre en correspondance le même texte dans deux langues différentes (corpus parallèles), en suivant donc la même approche que pour mettre en correspondance un signal de parole et sa transcription. Cette approche statistique a apporté des progrès importants qui conduisent à présent à la réalisation de systèmes de traduction dits hybrides, fondés à la fois sur des approches statistiques et sur des connaissances linguistiques.

5 META-NET White Paper Series, 2011

À présent, le défi est de traiter les informations liées au sens, aux niveaux sémantiques et pragmatiques, afin d'établir un dialogue naturel entre l'humain et la machine, ou de donner à la machine la capacité de participer à une communication entre humains. Pour ce faire, sont à prendre en compte d'autres modalités de communication (communication multimodale, traitement des documents multimédias), ainsi que le traitement des informations paralinguistiques (prosodie, expression des émotions, analyse des opinions et des sentiments).

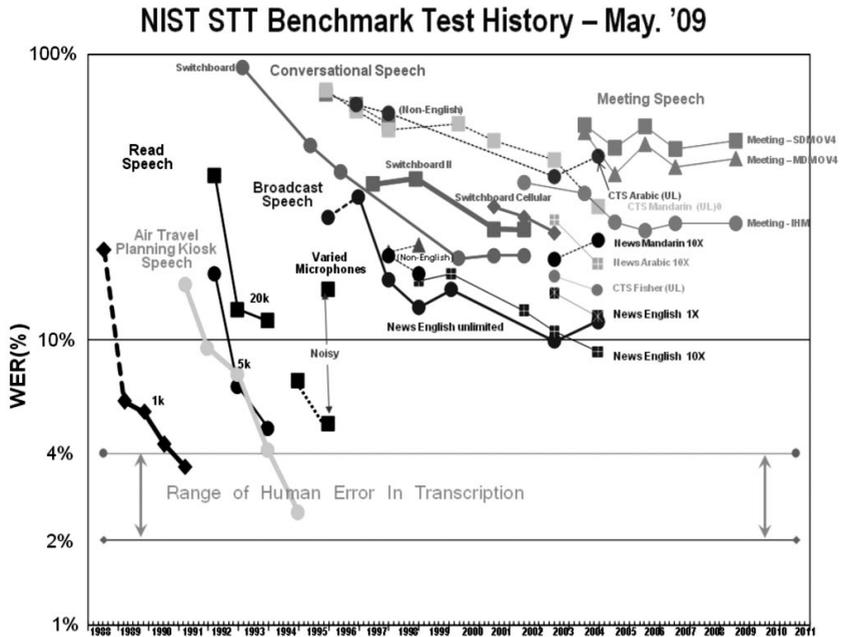
RESSOURCES LINGUISTIQUES ET ÉVALUATION

Il est crucial pour mener des recherches visant à développer les technologies de la langue de disposer d'un socle qui comporte à la fois des ressources linguistiques et des méthodes d'évaluation des technologies développées.

En ce qui concerne les ressources linguistiques, les données (corpus, lexiques, dictionnaires, bases terminologiques, etc.) sont à la fois nécessaires pour les recherches en linguistique et pour l'apprentissage des systèmes de traitement automatique de la langue qui fonctionnent, dans la plupart des cas, à partir de méthodes statistiques. Plus grande sera la quantité des données, meilleur sera le modèle statistique et meilleures seront donc les performances du système. L'interopérabilité des ressources linguistiques nous invite aussi à réfléchir aux normes à mettre en place pour organiser, parcourir et transmettre ces données.

Il est également nécessaire de disposer de moyens d'évaluation de ces technologies afin de pouvoir comparer les performances des systèmes selon un protocole commun, avec des données de test communes, dans le cadre de campagnes d'évaluation. Cela permet de comparer des approches différentes et d'avoir un indicateur de la qualité des recherches et de l'avancée technologique. À l'heure actuelle, on parle d'ailleurs de « coopération » – mélange de compétition et de coopération internationale – et cela est devenu un véritable moyen d'accompagnement de la recherche technologique. La Defense Advanced Research Project Agency (DARPA), du Département de la Défense des États-Unis, a été l'initiateur de cette démarche au milieu des années 80, à travers le National Institute of Standards and Technology (NIST) [MARIANI 1995].

Historique de la reconnaissance vocale depuis 1987 selon les évaluations du NIST⁶.



Ce tableau présente les progrès au fil des ans de la reconnaissance vocale, à travers les campagnes d'évaluation internationales menées par le NIST. Apparaissent sur ce tableau les meilleures performances obtenues cette année là, en termes de taux d'erreur sur les mots (WER) dans une échelle logarithmique, l'effort pour passer de 100 % d'erreurs (le système ne reconnaît aucun mot) à 10 % étant comparable à celui nécessaire pour passer de 10 % d'erreurs à 1 %. Les tâches à accomplir deviennent de plus en plus difficiles au fil des ans : commande vocale avec un langage artificiel de 1000 mots, puis dictée vocale (20 000 mots), transcription d'émissions de radio/TV (anglais, arabe et mandarin), conversations téléphoniques (également en anglais, arabe et mandarin), transcriptions de réunions... Et doivent se réaliser avec des conditions variables (temps réel ou non, différentes qualités de prise de son). On voit que pour certaines tâches, les performances des systèmes rejoignent celles d'un auditeur humain, rendant ces systèmes opérationnels et commercialisables (c'est le cas

⁶ <http://itl.nist.gov/iad/mig/publications/ASRhistory/index.html>

des langages de commande tels ceux actuellement disponibles sur de nombreux téléphones mobiles). Par contre, on voit bien que pour des tâches plus complexes, les performances s'améliorent plus lentement, justifiant la poursuite de l'effort de recherche. La connaissance de ces performances permet de déterminer la faisabilité d'une application en fonction du niveau de qualité qu'elle nécessite. Ainsi, un système de recherche d'information dans les données audiovisuelles ne nécessite pas des taux de transcription de parole très élevés contrairement à des systèmes de dialogue vocal.

Une même approche a été utilisée pour suivre les progrès en matière de traduction automatique, en utilisant la métrique BLEU proposée en 2000 [PAPINENI ET AL. 2001], alors que les recherches en traduction automatique avaient été conduites pendant une cinquantaine d'années sans songer à mesurer la qualité des résultats pour guider les travaux. Cette mesure donne une comparaison encore rudimentaire entre le résultat du système et les traductions de traducteurs humains.

Performances des systèmes de traduction automatique sur 22 langues officielles de l'UE [KOEHN ET AL. 2009]

Translating between all EU-27 languages 

	Target Language																					
	en	bg	de	cs	da	el	es	et	fi	fr	hu	it	lt	lv	mt	nl	pl	pt	ro	sk	sl	sv
en	-	40.5	46.8	52.6	50.0	41.0	55.2	34.8	38.6	50.1	37.2	50.4	39.6	43.4	39.8	52.3	49.2	55.0	49.0	44.7	50.7	52.0
bg	61.3	-	38.7	39.4	39.6	34.5	46.9	25.5	26.7	42.4	22.0	43.5	29.3	29.1	25.9	44.9	35.1	45.9	36.8	34.1	34.1	39.9
de	53.6	26.3	-	35.4	43.1	32.8	47.1	26.7	29.5	39.4	27.6	42.7	27.6	30.3	19.8	50.2	30.2	44.1	30.7	29.4	31.4	41.2
cs	58.4	32.0	42.6	-	43.6	34.6	48.9	30.7	30.5	41.6	27.4	44.3	34.5	35.8	26.3	46.5	39.2	45.7	36.5	43.6	41.3	42.9
da	57.6	28.7	44.1	35.7	-	34.3	47.5	27.8	31.6	41.3	24.2	43.8	29.7	32.9	21.1	48.5	34.3	45.4	33.9	33.0	36.2	47.2
el	59.5	32.4	43.1	37.7	44.5	-	54.0	26.5	29.0	48.3	23.7	49.6	29.0	32.6	23.8	48.9	34.2	52.5	37.2	33.1	36.3	43.3
es	60.0	31.1	42.7	37.5	44.4	39.4	-	25.4	28.5	51.3	24.0	51.7	26.8	30.5	24.6	48.8	33.9	57.3	38.1	31.7	33.9	43.7
et	52.0	24.6	37.3	35.2	37.8	28.2	40.4	-	37.7	33.4	30.9	37.0	35.0	36.9	20.5	41.3	32.0	37.8	28.0	30.6	32.9	37.3
fi	49.3	23.2	36.0	32.0	37.9	27.2	39.7	34.9	-	29.5	27.2	36.6	30.5	32.5	19.4	40.6	28.8	37.5	26.5	27.3	28.2	37.6
fr	64.0	34.5	45.1	39.5	47.4	42.8	60.9	26.7	30.0	-	25.5	56.1	28.3	31.9	25.3	51.6	35.7	61.0	43.8	33.1	35.6	45.8
hu	48.0	24.7	34.3	30.0	33.0	25.5	34.1	29.6	29.4	30.7	-	33.5	29.6	31.9	18.1	36.1	29.8	34.2	25.7	25.6	28.2	30.5
it	61.0	32.1	44.3	38.9	45.8	40.6	26.9	25.0	29.7	52.7	24.2	-	29.4	32.6	24.6	50.5	35.2	56.5	39.3	32.5	34.7	44.3
lt	51.8	27.6	33.9	37.0	36.8	26.5	21.1	34.2	32.0	34.4	28.5	36.8	-	40.1	22.2	38.1	31.6	31.6	29.3	31.8	35.3	35.3
lv	54.0	29.1	35.0	37.8	38.5	29.7	8.0	34.2	32.4	35.6	29.3	38.9	38.4	-	23.3	41.5	34.4	39.6	31.0	33.3	37.1	38.0
mt	72.1	32.2	37.2	37.9	38.9	33.7	48.7	26.9	25.8	42.4	22.4	43.7	30.2	33.2	-	44.0	37.1	45.9	38.9	35.8	40.0	41.6
nl	56.9	29.3	46.9	37.0	45.4	35.3	49.7	27.5	29.8	43.4	25.3	44.5	28.6	31.7	22.0	-	32.0	47.7	33.0	30.1	34.6	43.6
pl	60.8	31.5	40.2	44.2	42.1	34.2	46.2	29.2	29.0	40.0	24.5	43.2	33.2	35.6	27.9	44.8	-	44.1	38.2	38.2	39.8	42.1
pt	60.7	31.4	42.9	38.4	42.8	40.2	60.7	26.4	29.2	53.2	23.8	52.8	28.0	31.5	24.8	49.3	34.5	-	39.4	32.1	34.4	43.9
ro	60.8	33.1	38.5	37.8	40.3	35.6	50.4	24.6	26.2	46.5	25.0	44.8	28.4	29.9	28.7	43.0	35.8	48.5	-	31.5	35.1	39.4
sk	60.8	32.6	39.4	48.1	41.0	33.3	46.2	29.8	28.4	39.4	27.4	41.8	33.8	36.7	28.5	44.4	39.0	43.3	35.3	-	42.6	41.8
sl	61.0	33.1	37.9	43.5	42.6	34.0	47.0	31.1	28.8	38.2	25.7	42.3	34.6	37.3	30.0	45.9	38.2	44.1	35.8	38.9	-	42.7
sv	58.5	26.9	41.0	35.6	46.6	33.3	46.6	27.4	30.9	38.9	22.7	42.0	28.2	31.0	23.7	45.6	32.2	44.2	32.7	31.3	33.5	-

(using the Acquis corpus) [from Koehn et al., 2009]

Ce tableau donne les meilleures performances obtenues pour 462 paires de langues officielles de l'Union Européenne (manque le gaélique irlandais), en termes de score BLEU (plus le score est élevé, meilleure est la traduction, un traducteur humain obtenant de l'ordre de 80). Les meilleurs résultats correspondent aux langues qui font l'objet d'importants efforts de recherche, dans le cadre de programmes coordonnés, et bénéficient de nombreux corpus parallèles (anglais, français, néerlandais, espagnol, allemand...); les plus mauvais aux langues qui n'ont pas fait l'objet d'efforts similaires, ou qui sont très différentes des autres langues (hongrois, maltais, finlandais...).

En faisant référence aux enjeux initiaux, nous pouvons donc dégager deux points nécessaires à une politique en matière de technologies de la langue: pouvoir disposer de ressources linguistiques dans chacune des langues, afin de veiller à la préservation des cultures (et donc des langues) et, en même temps, pour permettre la communication entre les humains, utiliser des technologies et des ressources interlingues (tels les corpus parallèles: même texte traduit dans plusieurs langues) pour chacune des paires de langues que l'on veut traiter.

Mais il s'agit là d'une dynamique complexe: il y a un intérêt à développer des technologies monolingues pour chacune des langues afin de mieux traiter les technologies interlingues. Cela permet de faciliter la coordination des efforts: normes d'échange des données et des outils, retours d'expérience, recueils de bonnes pratiques. C'est aussi une nécessité pour des applications comme la traduction vocale (reconnaissance de la parole dans la langue source, traduction, puis synthèse vocale dans la langue cible) ou la recherche d'information interlingue (lorsqu'on veut produire un résumé de l'information qui a été trouvée, quelle que soit la langue dans laquelle elle a été codée), ou, plus généralement, pour la localisation de documents, qui nécessite à la fois des technologies interlingues (traduction...) et monolingues (correcteurs orthographiques...). Et cela permet également de faciliter un effort partagé entre les différents laboratoires de par le monde, qui travaillent souvent principalement sur leur langue nationale.

FRACTURE NUMÉRIQUE ET COUVERTURE LINGUISTIQUE

On note actuellement une situation à deux vitesses et une « fracture numérique », entre les langues pour lesquelles il existe des technologies et les autres. Cela est lié au « poids des langues »⁷ [GASQUET-CYRUS, PETITJEAN 2009]. Il faut noter que 95% des langues ne sont parlées que par 6 % de la population mondiale. Certains linguistes pensent que 90 % des langues auront disparu dans un siècle. On peut donc référencer les langues selon qu'elles sont bien, peu, ou pas « dotées » du tout, voire uniquement de tradition orale lorsqu'elles n'ont pas de système d'écriture, selon les données et les systèmes de traitement automatique qui existent pour ces langues. La disponibilité de données est cruciale pour développer des systèmes utilisables, souvent fondés sur des approches statistiques. La traduction automatique nécessite ainsi des corpus parallèles, dont le nombre est réduit⁸. On essaye donc de palier ce manque en développant des méthodes utilisant des corpus parallèles bruités, des corpus comparables (textes traitant d'un même sujet dans des langues différentes) ou quasi-comparables, qui sont plus aisément disponibles, notamment avec l'extension du web.

Pour résoudre cette fracture numérique, comment prendre en considération les langues dites « minoritaires », les langues régionales, les langues parlées par les migrants, les accents étrangers ou régionaux? Qui prend en charge le coût que cela représente, lorsque ces langues ne présentent pas un intérêt économique ou politique, et ne sont pas liées à des conflits armés ou des catastrophes naturelles qui justifient de les traiter? Comment assurer que les citoyens d'une communauté d'États puissent communiquer entre eux? Comment réduire les risques de conflits et de crises en permettant les échanges entre les peuples? Cela constitue de nos jours un enjeu politique et sociétal majeur, qui fait l'objet de nombreux débats. Ainsi, le Forum International de Bamako, organisé en janvier 2009 dans la poursuite des conclusions des Sommets Mondiaux pour la Société de l'Information de Genève (2003) et Tunis (2005), a conclu sur un engagement à promouvoir une utilisation éthique de l'information

7 Voir dans cet ouvrage: Daniel Prado, *Présence des langues dans le monde réel et le cyberspace*.

8 Voir dans cet ouvrage: Mikami Yoshiki & Shigeaki Kodama, *Mesurer la diversité linguistique sur le web*.

dans sa dimension linguistique, à permettre l'éducation dans la langue maternelle et à veiller à l'existence d'un cyberspace multilingue, tant dans les contenus disponibles sur la toile que dans les technologies permettant d'y accéder.

LES EFFORTS DE RECHERCHE DANS LE DOMAINE

Pour parvenir à produire les ressources linguistiques et les technologies nécessaires pour traiter le multilinguisme, différentes initiatives peuvent être recensées :

- celles de grandes compagnies comme Google ou Microsoft ;
- les programmes nationaux menés dans certains pays, avec différents objectifs: traiter un multilinguisme interne (TDIL en Inde, NHN en Afrique du Sud), comprendre des langues étrangères pour des raisons géopolitiques (GALE ou EARS aux États-Unis, financés par le Département de la Défense – DARPA), veiller à assurer l'usage et la promotion d'une langue nationale ou transnationale (TechnoLangue pour le Français, STEVIN pour le Néerlandais), ou à garder une place dans la compétition économique et culturelle (Quaero en France) ;
- les efforts de soutien à la R&D des programmes de la Commission Européenne ;
- les actions internationales visant à mettre en réseau les acteurs du domaine, pour mieux coordonner les activités et mieux partager les ressources ((Oriental) Cocosda, Clarin, FLaReNet, META-NET...) et la mise en place d'agences de distribution de ressources linguistiques, comme LDC aux États-Unis ou ELRA en Europe.

Ces différentes initiatives pour traiter le multilinguisme ont leurs avantages et leurs défauts : pérennité, liens avec la communauté scientifique, liens avec les applications, contrôle de qualité...

Industriels de l'informatique

Il faut tout d'abord souligner que les grandes sociétés américaines du secteur de l'information font un effort important pour le multilinguisme et l'interlinguisme. Ainsi, les moteurs de recherche de *Google* fonctionnent dans 145 langues (nationales et régionales), et *Google* a réalisé des outils

«gratuits» de traduction automatique et de recherche d'information interlingue en ligne : en avril 2011, 52 langues (dont le catalan et le galicien) et 2652 paires de langues étaient disponibles sur internet, 58 langues et 3306 paires de langues sur Smartphone (dont 16 langues avec entrée vocale, et 24 langues avec sortie vocale). La bibliothèque *Google Book Search* contenait pour sa part 7 millions de documents en 44 langues, et en décembre 2010 Google diffusait des statistiques sur l'évolution des langues, à partir d'un corpus de 500 milliards de mots (dont 361 milliards en anglais et 45 milliards de mots en français et en espagnol). *Microsoft* a également veillé à proposer le correcteur orthographique de Word en 126 langues (233 si l'on considère les variantes régionales) et un correcteur grammatical dans 6 langues (61 si l'on considère les variantes).

Des programmes nationaux portant sur les technologies de la langue pour aider le multilinguisme : TDIL en Inde, NHN en Afrique du Sud

De grands programmes ont été lancés dans le cadre de politiques publiques. Le programme TDIL⁹ (*Technology Development for Indian Languages*) est un important programme, qui constitue une des dix priorités du programme national indien sur l'informatisation de la société. Il s'agit de traiter l'anglais (indien) et les dix-huit langues indiennes officielles¹⁰, pour plusieurs technologies de la langue : traduction automatique, synthèse à partir du texte, reconnaissance vocale, moteurs de recherche, reconnaissance de caractères, correcteurs orthographiques, production de ressources linguistiques, tout cela pour l'ensemble de ces dix-neuf langues. Un programme comparable (NHN¹¹ *National Human Language Network*) se met en place en Afrique du Sud pour le traitement automatique des 11 langues nationales¹².

9 <http://tdil.mit.gov.in/>

10 assamese, bengali, gujrati, hindi, kannada, kashmiri, konkani, malayalam, manipuri, marathi, napali, oriya, punjabi, sanskrit, sindhi, tamil, telegu, urdu.

11 <http://www.meraka.org.za/nhn>

12 afrikaans, (south african) english, isiNdebele, isiXhosa, isiZulu, sepedi, sesotho, setswana, siSwati, tshivenda, xitsonga.

TechnoLangue : un programme pour le traitement du Français

En France, TechnoLangue [CHAUDIRON, MARIANI 2006]¹³, mené de 2002 à 2006, est un programme national qui, d'une part, a consisté à produire des ressources linguistiques (dictionnaires monolingues, spécialisés et bilingues, lexiques, corpus, bases terminologiques, outils de traitement...) et, d'autre part, à mener des campagnes d'évaluation pour le traitement de la langue écrite et de la langue parlée. Différentes campagnes ont été organisées pour le traitement du français, sur l'analyse syntaxique, l'extraction automatique de terminologie, les moteurs de recherche d'information de type Réponse aux Questions, la synthèse vocale à partir du texte, le dialogue oral et, plus particulièrement, sur la transcription de parole (pour l'indexation automatique d'émissions de radio ou de télévision). Dans ce cadre-là, un corpus important a été produit comportant 1600 heures de parole, dont 100 heures transcrites, ce qui représente un million de mots et 350 locuteurs enregistrés. Il n'existait pas jusqu'alors de corpus de cette taille pour d'autres langues que l'anglais américain; il était donc important d'en constituer un pour la langue française, et il reste important de le faire pour la plupart des langues dans le monde si l'on veut développer des systèmes qui les traitent automatiquement avec une qualité suffisante. TechnoLangue a également mené deux campagnes d'évaluation sur des technologies interlingues. L'une sur l'alignement de textes parallèles, d'une part entre le français et l'anglais, l'allemand, l'italien et l'espagnol; et d'autre part, entre des langues ayant des graphies différentes: entre le français et l'arabe, le mandarin, le grec, le japonais, le persan et le russe. Enfin, une campagne d'évaluation a été menée sur la traduction automatique entre l'anglais et le français et entre l'arabe et le français, incluant une réflexion sur les métriques d'évaluation en traduction automatique.

13 <http://www.technolangue.net/>

QUAERO : un programme français pour le traitement des documents multilingues et multimédias

Le programme QUAERO¹⁴ a été lancé en France en mai 2008. Il porte sur le traitement des documents multilingues et multimédias. Le programme est structuré autour du développement d'une trentaine de technologies sur différents médias (texte, parole, image, vidéo, musique...) qui répondent aux demandes d'un ensemble de cinq applications différentes (Numérisation et enrichissement de contenus; Analyse des médias; Vidéo personnalisée; Moteurs de recherche; Portails de communication). Il est fondé sur l'utilisation de corpus et d'évaluation. Il est prévu de traiter plus d'une vingtaine de langues. Ce programme, regroupant 26 partenaires publics et privés, est doté d'un budget de 200 millions d'euros sur cinq ans (2008-2013), avec 100 millions d'euros d'aides publiques apportées par l'agence Oseo. Les premiers résultats ont abouti par exemple au moteur de recherche Voxlead sur les données audiovisuelles¹⁵ (radio, télévision, vidéos en ligne...) dans six langues (français, anglais, espagnol, arabe, mandarin et russe) développé par Exalead, à un agrégateur d'actualités plurimédias (textes, radios, télévisions) développé par Orange ou à un système de lecture de livre électronique par Jouve.

Actions de l'Union Européenne

L'Union Européenne a bénéficié de 2007 à 2010 d'un commissaire spécifiquement pour le multilinguisme¹⁶, qui a mis en place un Groupe de Haut Niveau sur le Multilinguisme, a produit un rapport¹⁷, et a présenté une communication au Parlement et au Conseil Européen en septembre 2008¹⁸. Dans le cadre de la présidence de l'Union Européenne, la France a organisé en septembre 2008 les États-Généraux du Multilinguisme à La Sorbonne (Paris), qui ont été suivis en novembre 2008 d'une résolution

14 <http://www.quaero.org>

15 <http://voxleadnews.labs.exalead.com/>

16 http://ec.europa.eu/commission_2004-2009/orban/index_fr.htm

17 http://ec.europa.eu/education/policies/lang/doc/multireport_en.pdf

18 <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/08/1340&format=HTML&aged=0&language=EN&guiLanguage=fr>

du Conseil des Ministres Européens sur le multilinguisme¹⁹, reprise ensuite par le Parlement Européen en mars 2009²⁰. L'idée d'un « *Single European Information Space* » a été mise en avant.

La Commission Européenne a soutenu plusieurs projets importants sur les technologies multilingues dans le cadre du 6^e Programme-Cadre de Recherche et Développement (CLEF, TC-STAR, CHIL...). En particulier, le projet intégré TC-STAR²¹ a porté sur la traduction vocale pour trois langues – l'anglais, l'espagnol et le chinois – dans le cadre d'une application d'interprétation automatique des interventions au Parlement Européen. Travailler dans ce cadre est très intéressant puisque les ressources existent au Parlement Européen : les interventions des députés dans leur langue, leur interprétation dans les différentes langues de l'hémicycle, leur transcription sous forme écrite, et la traduction de cette transcription écrite dans les différentes langues officielles. Ainsi, ces données permettent l'apprentissage par les moteurs des systèmes d'interprétation automatique, incluant reconnaissance dans la langue source, traduction de la langue source à la langue cible, et synthèse dans la langue cible, l'application faisant donc appel à des technologies monolingues et interlingues. Une démonstration d'un système pour le couple de langue anglais-espagnol est disponible en ligne²². TC-STAR a également produit et diffusé un rapport en cinq langues sur les technologies de la langue en Europe [LAZZARI, STEINBISS 2006]²³.

Dans le 7^e PCRD (2007- 2013), ce secteur est traité en particulier par l'Unité « Technologies du langage, Traduction Automatique ». Outre le soutien aux projets de R&D, une infrastructure et deux réseaux ont été mis en place : CLARIN (*Common Language Resources and Technology Infrastructure*)²⁴, FLARENET (*Fostering Language Resources Network*)²⁵ et META-NET (*Multilingual Europe Technology Alliance*)²⁶.

19 <http://eur-lex.europa.eu/Notice.do?mode=dbl&lang=en&ihtmlang=en&lng1=en,fr&lng2=bg,cs,da,de,el,en,es,et,fi,fr,hu,it,lt,lv,mt,nl,pl,pt,ro,sk,sl,sv,&val=471776:cs&page=>

20 <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2009-0162+0+DOC+XML+V0//FR&language=FR>

21 <http://www.tc-star.org/>

22 Voir « Demo JM.asf » sur le site web http://audiosurf.org/demo_video/

23 http://www.tc-star.org/pubblicazioni/D17_HLT_ENG.pdf

24 <http://www.clarin.eu/>

25 <http://www.flarenet.eu/>

26 <http://www.meta-net.eu/>

CLARIN est une infrastructure soutenue dans le cadre du programme ESFRI (*European Strategy Forum on Research Infrastructure*) de la Commission Européenne. Son objectif est la distribution de ressources linguistiques et d'outils pour les Sciences Humaines et Sociales.

FLaReNet est un Réseau Thématique soutenu dans le cadre du programme Européen e-Content, avec un budget de 0.9 millions d'euros sur 3 ans (2008-2011). Son objectif est de servir de *Think Tank* pour la promotion des ressources linguistiques dans les programmes Européens.

Le réseau d'Excellence META-NET a été mis en place dans le cadre du projet T4ME (*Technologies for a Multilingual Europe*). Ce projet d'un budget de 6 millions d'euros sur une durée de 3 ans (2010-2013) est structuré en trois volets :

- Pousser les frontières de la recherche en traduction automatique ;
- Mettre en place une infrastructure ouverte de ressources (META-SHARE), incluant la production, l'annotation, la standardisation, la validation et la distribution de ressources, et l'évaluation de technologies ;
- Mener une réflexion sur les technologies multilingues dans la perspective d'un *Strategic Research Agenda* pour le 8^e PCRD.

PERSPECTIVE EUROPÉENNE ET INTERNATIONALE

Les résolutions des instances européennes demandent de consacrer un effort très important pour traiter l'ensemble des langues européennes, nationales et régionales. Cependant, si l'on considère le nombre de langues ou de paires de langues, et qu'on le multiplie par le nombre de technologies, on voit que la taille de l'effort est probablement trop importante pour la Commission Européenne seule. Il y aurait donc intérêt à partager cet effort entre les États-Membres ou les Régions et la Commission Européenne, en parfaite harmonie avec le « principe de subsidiarité ».

Les technologies de la langue sont bien adaptées à un effort commun. La Commission Européenne aurait en priorité la responsabilité de veiller à assurer la coordination du programme (gestion, mise à disposition de normes et standards, évaluation de technologies, communication,...) et de développer des technologies génériques autour du traitement des langues. Chaque État-Membre pourrait en priorité à assurer la couverture de sa, ou de ses langues quand il y en a plusieurs : produire les ressources

linguistiques, essentielles pour développer les systèmes (corpus, lexiques, dictionnaire), et développer ou adapter les technologies de la langue aux spécificités de sa, ou de ses langues. Ce modèle serait facilement adaptable à un effort international, en conjuguant les efforts des pays partenaires et des organismes internationaux.

Malheureusement, jusqu'à présent le thème des technologies de la langue a été considéré comme un domaine de recherche parmi d'autres dans les programmes européens, et non comme une donnée essentielle de la construction européenne, qui nécessite un effort à la hauteur de ces enjeux. Cette faiblesse est d'autant plus dangereuse que la vivacité de l'Union Européenne a besoin de multiplier les échanges économiques, informationnels et culturels entre pays, et de s'adresser aux citoyens de chacun d'entre eux, avec ses spécificités, en les aidant également à communiquer. Espérons que la prise de conscience politique des enjeux du multilinguisme verra la recherche dans les technologies de la langue bénéficier de l'attention nécessaire dans les Programmes-Cadres à venir.

CONCLUSION

Les technologies de la langue sont un outil majeur pour faciliter le multilinguisme en Europe comme dans le reste du monde. Pour y parvenir, il conviendrait de coordonner les efforts des États, voire des Régions, et des organismes internationaux (Commission Européenne, Organisation des Nations Unies, Unesco, Union Africaine...), en y associant les industriels et les laboratoires de recherche publics. Il faudrait veiller à produire, pour chacune des langues, les ressources linguistiques nécessaires, et organiser l'effort de recherche de manière ouverte en le fondant sur l'interopérabilité et l'évaluation comparative et objective des technologies.

On pourrait alors ajouter un clin d'œil à la phrase célèbre d'Umberto Eco, en disant : « *La langue universelle, c'est la traduction... avec le soutien de la technologie!* »

BIBLIOGRAPHIE

[CENCIONI, ROSSI 2008] R. Cencioni, K. Rossi. Language based Interaction, *EC-ICT Conference*, Lyon, 26 Novembre 2008.

[CHAUDIRON, MARIANI 2006] S. Chaudiron, J. Mariani. Techno-langue: The French National Initiative for Human Language Technologies (HLT), *Proceedings LREC'06*, Genoa, Italy, May 2006.

[ECO 1993] U. Eco. *La langue de l'Europe, c'est la traduction*, Assises de la traduction littéraire, Arles, 1993.

[GASQUET-CYRUS, PETITJEAN EDS 2006] M. Gasquet-Cyrus, C. Petitjean eds. *Le poids des langues*, L'Harmattan, 2009.

[KOEHN, BIRCH, STEINBERGER 2009] Ph. Koehn, A. Birch and R. Steinberger. 462 Machine Translation Systems for Europe, *Machine Translation Summit XII*, p. 65-72, 2009.

[LAZZARI, STEINBISS 2006] G. Lazzari, V. Steinbiss. *Human Language Technologies for Europe*, TC-Star Report, April 2006.

[MARIANI 1995] J. Mariani, ed. *Evaluation chapter in Survey of the State of the Art in Human Language Technology*, R. A. Cole, J. Mariani, H. Uszkoreit, N. Varile, A. Zaenen, A. Zampolli, V. Zue eds., Cambridge University Press, 1995.

[PAPINENI, ROUKOS, WARD, ZHU] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu. BLEU: A Method for Automatic Evaluation of Machine Translation. In: *Proceedings of the 40th Annual Meeting of ACL*, Philadelphia, PA.

L'USAGE DE FACEBOOK CHEZ LES ÉTON DU CAMEROUN

Ouvert au public en 2006, Facebook compte à ce jour des centaines de millions d'utilisateurs à travers le monde, qui y utilisent plusieurs dizaines de langues. En prenant l'exemple des groupes Facebook animés par et pour des membres du groupe ethnique Éton (Centre-Cameroun, approximativement 250 000 locuteurs), nous pouvons observer comment l'utilisation des technologies de communication récentes permet de nouvelles formes d'affirmation des cultures traditionnelles.

Article original en français.



VASSILI RIVRON est anthropologue, maître de conférences à l'Université de Caen-Basse Normandie. Il est rattaché au CERReV (UCBN) et chercheur associé à l'École des Hautes Études en Sciences Sociales et au CSU (CNRS).

VASSILI RIVRON

L'USAGE DE
FACEBOOK
CHEZ LES ÉTON
DU CAMEROUN

Ouvert au public en 2006, le site Facebook compte à ce jour plusieurs centaines de millions d'utilisateurs à travers le monde. L'interface du site est disponible en plusieurs dizaines de langues (y compris des langues régionales comme le basque, ou artificielles comme l'espéranto), faisant appel à plusieurs systèmes d'écriture, alphabétiques ou non. Ce «réseau social» permet des usages très diversifiés, parmi lesquels on retrouve des projections sur le web de réseaux, de pratiques et de références culturelles préexistants, mais également des innovations culturelles proprement dites.

En prenant l'exemple de plusieurs «groupes Facebook» animés par et pour des membres du groupe «ethnique» Éton (Centre-Cameroun, approximativement 250 000 locuteurs), nous pouvons observer comment l'utilisation des technologies de communication récentes permet de nouvelles formes d'affirmation des cultures traditionnelles. Parmi celles-ci, nous avons pu constater une extension de l'usage d'une langue dite «maternelle» hors des contextes et registres habituels, associé au développement de la pratique de son écriture.

Dans un pays où l'on recense deux langues officielles (français et anglais) et plus de deux cents langues locales, la langue éton¹ semblait comme nombre d'autres, devoir rester confinée à des contextes géographiques ou sociaux déterminés. D'une part le département de la Lékié et d'autre part, pour les migrants vers la ville ou l'étranger, l'espace domestique, les cercles familiaux, les hiérarchies traditionnelles, et ceci principalement, dans le cadre d'une communication interpersonnelle (y compris téléphonique) ou du moins, en présence des interlocuteurs (de type discours).

1 Langue tonale bantoue, au Nord du complexe linguistique beti / bulu / fang.

Quelques langues camerounaises sont écrites ou l'ont été (y compris par des systèmes graphiques spécifiques). Certaines ont été enseignées pendant la période du protectorat allemand (1884-1922), avant que cela ne soit aboli sous la colonisation française. La perspective de ce type d'enseignement revient périodiquement à l'ordre du jour depuis l'Indépendance (1960) mais est contrée, dans l'imaginaire républicain, par la crainte de dérives tribalistes. De ce fait, l'éton demeure à ce jour principalement oral et son écriture usuelle n'est toujours pas codifiée. En réalité, cette écriture semble faire principalement l'objet de productions savantes : transcriptions ethnographiques, études linguistiques, publication de textes bilingues dans des recueils de folklore ou de compilations de littératures orales... Elle sert par ailleurs de ressource mnémotechnique à ses locuteurs dans le cadre d'annotations personnelles (mémoires, dictons, notes) ou dans des passages de connivence ou d'érudition au sein de correspondances qui sont principalement rédigées dans une autre langue². L'absence d'enseignement et les principes républicains hérités de la France expliquent l'impossibilité d'un lectorat régulier en éton (presse, ouvrages, administration, politique).

Dès le début de nos observations (2004), nous avons pu constater une extension de l'éton écrit sur le Web, et ce principalement dans deux contextes. D'une part un certain nombre d'initiatives (individuelles souvent) ont mené à la création de sites culturels « patrimoniaux » (sur le folklore, la culture, la langue, l'histoire du groupe ethnique ou de sa région). L'apparition de l'éton y prend la forme de lexiques fragmentaires, ou encore de recensements de dictons, par exemple. Il s'agit rarement dans ces sites de s'adresser directement à un public en éton écrit, mais plutôt de transcrire l'oral ou d'enseigner l'oral. L'apparition de dialogues écrits en éton s'est faite logiquement du côté des blogs et forums. Nous avons pu le constater notamment dans les commentaires aux vidéos d'artistes locaux sur YouTube ou dans les réponses aux articles de la presse camerounaise sur le Web. Mais dans un cas comme dans l'autre, l'usage de la langue dépasse rarement deux phrases sans une traduction et se résume souvent à de l'illustration, de l'interpellation, des traits d'esprit ou des propos de connivence (excluant les non-locuteurs).

2 Voir à propos des techniques et langues de la correspondance avec la diaspora : Sayad, Abdelmalek, « Du message oral au message sur cassette : la communication avec l'absent », *Actes de la recherche en sciences sociales*, n°59, 1985, pp. 61-72.

L'apparition des « réseaux sociaux » (ici Facebook) a suscité le développement d'une grande diversité de communautés (« groupes ») où se recoupent généralement des entités individuelles (avec photo et données personnelles) par des procédures de cooptation (demandes d'« amitié ») et qui donnent alors accès aux contenus des « profils » ou des « groupes » (photos, textes, vidéos, jeux, etc.) que l'on peut partager de façon paramétrable. Parmi les groupes Facebook intéressants pour notre propos, nous pouvons citer « Les Étons » (réfèrent ethnique, 53 membres), « Fils et filles de la Lékié » (réfèrent territorial, 220 membres) et « Ongola - Culture Fang-Bulu-Beti » (réfèrent culturel élargi : 1445 membres). Outre l'idée de transposer différents aspects de la structure sociale et des pratiques culturelles des groupes sur une ressource électronique, nous y trouvons également des formes d'écrit qui évoluent par rapport à ce que nous avons observé initialement. Parmi les nombreuses thématiques abordées dans ces « groupes Facebook », une partie importante est consacrée à des questions génériques touchant à la culture « traditionnelle », le mariage, la parenté, les initiations, les dictons. La question de la langue fait rarement l'objet d'un traitement spécifique.

Des échanges durables se nouent sur les réseaux sociaux, utilisant différents registres. Le français ou l'anglais restent clairement dominants, que cela soit pour les propos génériques (description, instructions du groupe) ou pour les commentaires et dialogues eux-mêmes. Mais au sein des messageries, « murs » et forums de discussion propres à chaque groupe, l'éton peut apparaître dès le titre de la rubrique. Il peut ensuite ouvrir le débat et se poursuivre (y compris dans d'autres langues).

Il semble y avoir moins de gêne à écrire l'éton dans l'« entre-soi » des groupes Facebook, l'intercompréhension étant postulée malgré l'absence de graphie usuelle codifiée et enseignée. L'inégalité des compétences linguistiques se manifeste clairement dans ces échanges, ainsi que la diversité des ressources mobilisées pour concevoir des solutions graphiques (dans un environnement francophone, anglophone, ewondophone³). Les efforts pour se faire comprendre par écrit dans cette langue tonale, comme pour déchiffrer les messages (en les ré-oraisant), sont visibles dans les échanges et lors des observations que nous avons pu faire en situation. Le plaisir

3 Ewondo : langue parlée dans la partie sud du Cameroun et notamment à la capitale, Yaoundé.

à le faire est lui aussi manifeste : cette expression relève de la complicité et de la fierté culturelle.

Ce qui est intéressant pour le chercheur, c'est que ce corpus textuel est complété par des « profils » qui donnent des indications précieuses sur les membres actifs du groupe. Loin d'être une émanation populaire, le réinvestissement et les déplacements des ressources tenues pour « traditionnelles » sur internet sont en grande partie le fait de lettrés souvent cosmopolites, si ce n'est expatriés. Nous constatons effectivement, dans les cas cités plus haut, que les fondateurs ou les modérateurs/animateurs de ces groupes étaient, respectivement : un blanc du pays, une camerounaise expatriée aux États-Unis et un intellectuel camerounais. Il n'en pouvait être autrement, déjà pour de simples raisons matérielles : Facebook ne peut pas être consulté sur les nombreux terminaux obsolètes que l'on trouve au Cameroun. Mais au-delà, et comme nous pouvons l'observer pour la codification de la culture populaire européenne par les folkloristes du XIX^e et XX^e siècles, le souci et la capacité à s'approprier des ressources informatiques et à transposer des ressources linguistiques d'une langue à l'autre, comme de l'oral vers l'écriture, sont socialement déterminées et consacrent le rôle déterminant des « élites cosmopolites » et en particulier de celles qu'on dit « passées de l'autre côté » (diaspora). Un détour vraisemblablement nécessaire pour assurer la vitalité d'une langue orale dans le contexte écrit des nouvelles technologies.

BIBLIOGRAPHIE

- [ABÉLÈS 2008] Abélès, Marc, *Anthropologie de la globalisation*, Paris, Payot, 2008.
- [AMSELLE 1999] Amselle, Jean-Loup et Mbokolo, Elikia (dir.), *Au cœur de l'ethnie*, Paris, La Découverte, 1999.
- [BOURDIEU 1994] Bourdieu, Pierre : « Esprits d'État – Genèse et structure du champ bureaucratique », In : *Raisons Pratiques*, Seuil, Paris, 1994, pp 99-135.
- [GOODY 1994] Goody, Jack P, *Entre l'oralité et l'écriture*, PUF, Paris, 1994.
- [GUICHARD 2003] Guichard, Eric, « Does the 'Digital Divide' Exist? », In : *Globalization and its new divides: malcontents, recipes, and reform* (dir. Paul van Seters, Bas de Gaay Fortman & Arie de Ruijter), Dutch University Press, Amsterdam, 2003.
- [GUYER 2000] Guyer, Jane I., « La tradition de l'invention en Afrique équatoriale », *Politique africaine*, n° 79, octobre 2000, pp.101-139.

[SAYAD 1985] Sayad, Abdelmalek, « Du message oral au message sur cassette : la communication avec l'absent », *Actes de la recherche en sciences sociales*, n°59, 1985, pp. 61-72.

[THIESSE 1999] Thiesse, Anne-Marie, *La création des identités nationales (Europe XVIII^e-XX^e siècle)*, Seuil (coll. Univers Historique), Paris, 1999.

[VAN VELDE 2006] Van Velde, Mark, *A description of Eton : phonology, morphology, basic syntax and lexicon*, thèse de doctorat, 2006.

PANN YU MON
& MADHUKARA PHATAK

MOTEURS DE RECHERCHE ET LANGUES ASIATIQUES

Bien qu'un grand nombre de moteurs de recherche soient accessibles dans les langues les plus utilisées au sein du monde numérique, ils ne fonctionnent plus lorsqu'ils doivent traiter des langues moins informatisées. Ces dernières années, le nombre de ressources non anglophones sur le web a augmenté rapidement, notamment dans les langues asiatiques. Cet article soulève les difficultés rencontrées par les moteurs de recherche dans cette situation.

Article original en anglais.
Traduit par Laurent Vannini.



PANN YU MON est titulaire d'un doctorat soutenu au département Management et ingénierie des systèmes d'information de l'Université de Technologie de Nagaoka, Japon. Ses champs de recherche sont l'indexation, l'archivage et les requêtes web.



MADHUKARA PHATAK est Ingénieur en Informatique diplômé de l'université JSSATE en Inde. Ses recherches portent sur l'informatique en nuages et les systèmes distribués.

PAN YU MON
& MADHUKARA PHATAK

MOTEURS DE
RECHERCHE
ET LANGUES
ASIATIQUES

Véritable cœur de la société de l'information, les moteurs de recherche jouent un rôle essentiel dans l'extraction de données issues d'un internet à la profondeur quasi insondable. Bien que plusieurs moteurs de recherche soient accessibles dans les langues dominantes, ils sont inefficaces lorsqu'entrent en scène des langues moins informatisées. Au cours des dernières années, le nombre de ressources non-anglophones sur internet n'a cessé d'augmenter et on estime que pour plus de 60% des utilisateurs du Web, l'anglais n'est pas la langue native. Quelle que soit la précision de ces données, il est évident désormais que les contenus et utilisateurs non anglophones ne doivent plus être ignorés.

Lorsque les moteurs de recherche dominants traitent des requêtes qui ne sont pas effectuées en anglais, ils ne font que du filtrage par motif – la séquence de symboles entrée par l'utilisateur correspond à un contenu identique présent quelque part dans un document sur le Web. Des méthodes plus sophistiquées, s'appuyant sur le traitement automatique du langage naturel¹, ne sont pas utilisées (comme par exemple la lemmatisation, l'extraction de mots vides ou encore l'analyse lexicale).

Pour répondre aux besoins de locuteurs de langues différentes, Google a développé plus de 136 interfaces linguistiques et propose environ 180 moteurs de recherche régionaux ou locaux. Seuls 20% d'entre eux sont dédiés aux langues asiatiques. Un faible nombre d'articles a été écrit sur les difficultés rencontrées dans le traitement des requêtes en langues occidentales; mais un plus petit nombre encore a analysé les requêtes en langues asiatiques. Notre principal objectif est ici de débattre des problèmes supplémentaires inhérents aux requêtes non anglophones, et de suggérer des outils et techniques pour améliorer la réponse des systèmes

1 Voir dans cet ouvrage: Joseph Mariani, *Les technologies de la langue en soutien au multilinguisme*.

de recherche existants. Nous étudions dans cet article les difficultés rencontrées par les moteurs de recherche lorsqu'ils traitent des requêtes en langues asiatiques. Nous donnerons des exemples en cinq langues ou familles de langues : les langues indiennes, malaisiennes, birmanes, indonésiennes et thaï.

INTRODUCTION

Les moteurs de recherche doivent parcourir des milliards de pages web afin d'indexer l'hypertexte, qui est en évolution permanente et contient des informations dans une grande variété de langues et de formats. La croissance du Web est exponentielle, et le nombre de pages indexables² approche désormais les cent milliards. Il est devenu de plus en plus difficile pour les moteurs de recherche de maintenir une indexation automatique actualisée et complète, ce qui a pour conséquence une dégradation de la précision et de la couverture des méthodes de traitement. Les utilisateurs ont souvent beaucoup de mal à chercher et trouver de l'information utile et de bonne qualité sur le Web par le biais de moteurs de recherche généralistes, et ce notamment lorsqu'ils sont en quête d'information sur un sujet pointu ou dans une langue autre que l'anglais. Un moteur de recherche ne peut pas être performant dans toutes les langues. C'est pour cette raison que les moteurs de recherche doivent être localisés dans les langues régionales. Un grand nombre de moteurs de recherche thématiques ou développés dans une langue singulière ont été construits dans diverses régions du monde pour rendre les recherches plus efficaces. Nous visons dans cet article à identifier les différents types de difficultés rencontrés par les moteurs de recherche dans le traitement des requêtes en langues asiatiques. Bien qu'il existe des outils logiciels complets permettant la création de moteurs de recherche, la plupart d'entre eux ne fonctionnent pas avec certaines langues européennes, asiatiques ou du Moyen-Orient.

Les modules principaux d'un moteur de recherche sont :

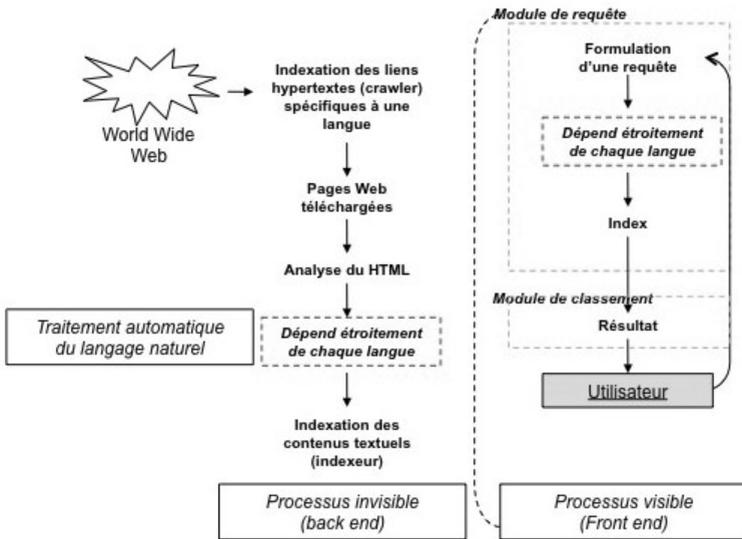
- un robot d'exploration, ou *crawler*
- un module de traitement automatique du langage naturel (TAL)
- un indexeur

2 Le Web surfacique (ou visible, ou indexable) est la partie du World Wide Web qui est accessible en ligne et indexable par les robots d'indexation des moteurs de recherche généralistes http://en.wikipedia.org/wiki/Surface_Web

- un module de requête
- un module algorithmique de classement

Le robot d'indexation est un petit programme qui parcourt le World Wide Web et télécharge des pages web. Sur la base d'un premier échantillon d'adresses web, ou d'URL pivots (*seed urls*), ce programme copie les pages et identifie d'autres URL à visiter. Afin d'implémenter un moteur de recherche spécifique à une langue, nous avons besoin d'une portion pertinente au regard de cette langue du World Wide Web.

Figure 1: Architecture d'un moteur de recherche



Afin de télécharger uniquement les contenus pertinents du World Wide Web des critères d'exploration spécifiques sont nécessaires. L'étape suivante, relativement simple, est celle de l'analyse du HTML pour extraire le contenu textuel à indexer. Cette étape sera réalisée par un processus de traitement automatique du langage naturel, qui est plus complexe pour les pages non anglophones. Le rayon d'action de ce module de traitement est variable selon la langue. Ces traitements comprennent les phases de

translittération³, de tokenization⁴, de racinisation⁵, de prétraitement des mots composés dans les requêtes, de suppression des mots vides (non significatifs) etc.

Le module suivant, l'indexation, extrait tous les mots de chaque page et établit une liste des URL où chaque mot est présent. Le résultat est en général une liste des URL des pages où apparaît un mot donné (fichier « inverse »). Cette étape inclut des tâches comme le transcodage, le découpage en mots, la racinisation et la suppression de mots vides.

Le module de requête a pour fonction de recevoir et de répondre aux recherches formulées par les utilisateurs. Le module de classement a pour rôle d'ordonner les résultats de telle manière que soient présentés en premiers les pages les plus pertinentes. Tous les modules composant le moteur de recherche ont la même fonction quelque soit la langue ciblée, à l'exception du module de traitement automatique de langage naturel qui dépend étroitement des caractéristiques propres à chaque langue.

À CHAQUE LANGUE SA DIFFICULTÉ SINGULIÈRE

Nous étudions dans cette partie les tâches liées au module de traitement automatique du langage naturel. Ces tâches peuvent varier selon la langue. Nous identifions ici au travers d'exemples les différents types de tâches selon les familles de langues.

Traitement de l'encodage

L'existence de différents encodages de caractères pour les textes composant les documents du web est une réalité qui devrait être prise en compte durant la phase d'indexation. Cela est particulièrement vrai dans le cas des langues asiatiques. Nous prenons ici les exemples de langues indiennes et

3 Opération qui consiste à substituer à chaque graphème d'un système d'écriture un graphème ou un groupe de graphèmes d'un autre système, indépendamment de la prononciation. Elle dépend donc du système d'écriture cible, mais pas de la langue http://fr.wikipedia.org/wiki/Transcription_et_translittération

4 Processus permettant de démarquer les différentes sections d'une chaîne de caractères dans l'analyse lexicale http://fr.wikipedia.org/wiki/Analyse_lexicale#Tokenization

5 Procédé qui vise à transformer les flexions en leur radical ou stemme. La racinisation cherche à rassembler les différentes variantes flexionnelle et dérivationnelle d'un mot autour d'un stemme <http://fr.wikipedia.org/wiki/Racinisation>

birmanes. Les langues indiennes ont toutes des encodages différents. Plus complexes encore, le birman du Myanmar présente plusieurs encodages pour une seule et même langue.

Les langues indiennes

Plus de 95 % du contenu en langues indiennes sur le Web ne peut faire l'objet d'une recherche en raison de l'utilisation d'encodages différents sur chaque page. La plupart de ces encodages sont incompatibles⁶ et nécessitent par conséquent la mise en œuvre de normes pour que les contenus encodés soient accessibles via un moteur de recherche. Les écritures indiennes, ou brahmiques, sont de nature phonétique. Elles sont composées de symboles de types voyelles et consonnes. Les consonnes deviennent une syllabe après l'ajout d'un son-voyelle. Pour compliquer la problématique, existent des « syllabes composées » appelées également ligatures. Par exemple, si nous nous arrêtons sur le « tri » de « triangle », on peut distinguer trois lettres correspondant à trois sons 'ta', 'ra' et 'yi'. Mais dans le cas des écritures indiennes, les trois lettres sont assemblées pour créer une seule syllabe composée, dénuée d'une structure linéaire, comme c'est le cas pour les langues latines⁷.

Un grand nombre de langues en Inde utilisent la même écriture, la Devanagari, ce qui rend la détection de la langue encore plus complexe. C'est le cas d'une requête portant sur « मधुकर », par exemple, qui a pour signification « pot de miel » en hindi et « fusil » en oriya.

Webkhoj est un moteur de recherche indien qui donne le choix aux utilisateurs d'effectuer des recherches parmi dix langues indiennes. Le moteur reconnaît l'hindi, le telugu, le tamoul, le malayalam, le marathi, le kannada, le bengali, le punjabi, le gujarati et l'oriya. Afin d'effectuer ses recherches sur des sites web écrits dans ces langues, Webkhoj translitère tous les encodages en un encodage standard (Unicode/UCS) et accepte les requêtes des utilisateurs dans le même encodage pour ensuite proposer les résultats de la recherche.

6 Voir dans cet ouvrage : Mikami Yoshiki & Shigeaki Kodama, *Mesurer la diversité linguistique sur le Web*.

7 Prasad Pingali, Jagadeesh Jagarlamudi, Vasudeva Varma, "WebKhoj: Indian language IR from Multiple Character Encodings", In: *WWW '06 Proceedings of the 15th international conference on World Wide Web*, 2006. <http://dl.acm.org/citation.cfm?doid=1135777.1135898>

Le birman/myanmar

Le myanmar⁸ utilise différents encodages. Lorsqu'un utilisateur opère une requête à l'aide de mots clés dans un encodage spécifique, le moteur n'effectue de recherche que dans les pages utilisant le même encodage. Certaines pages utilisant d'autres encodages peuvent ainsi ne pas être prises en compte. Par conséquent, certaines pages pertinentes peuvent être exclues du simple fait de l'utilisation d'encodages différents. Trois principaux ensembles d'encodages alternatifs ont été implémentés pour les pages web en birman par différents groupes de personnes.

Les encodages graphiques: en réalité, ces encodages se font passer pour des jeux de caractères anglais (techniquement, des jeux de caractères Windows-1252 ou iso-Latin-1) et substituent les glyphes (représentation graphique d'un caractère) du birman aux glyphes romains pour représenter les caractères de la langue locale.

Les encodages suivant partiellement les jeux de caractères et normes ucs-Unicode: ces types d'encodages ont des tables de caractères différentes dont aucune ne suit exactement les jeux de caractères de l'ucs (*Universal Coded-character Set*) ou les normes de l'Unicode⁹. Ils ne font que suivre partiellement ces standards, mais ne sont pas encore pris en charge par Microsoft ou d'autres producteurs de logiciels majeurs.

Les encodages ucs-Unicode: ces jeux de caractères se réfèrent non seulement aux tables et glyphes Unicode, mais également aux logiques et règles du format *OpenType* (OTL).

De surcroît, certaines pages web en birman/myanmar sont créées en utilisant ce qui est appelé le format *Mixture Encoding Style*. C'est le mélange de caractères issues des jeux et normes ucs-Unicode et d'entités HTML, comme `ံ`; `ြ`; `္`; `ိ`; `္`; `္`; (qui apparaîtra à l'écran comme: သံလွင်ဒေဝိမ်မက်). Les contenus sont codés en système décimal. Pour ce type de pages web, l'entité HTML devrait être convertie en codes ucs-Unicode en passant du système décimal au système hexadécimal. Certains logiciels de publication web encodent systématiquement les mots birmans dans le format

⁸ Dans un contexte linguistique, le terme birman évoque plutôt la langue parlée, tandis que celui de myanmar désigne la langue écrite [http://fr.wikipedia.org/wiki/Birman_\(langue\)](http://fr.wikipedia.org/wiki/Birman_(langue))

⁹ Voir dans cet ouvrage: Stéphane Bortzmeyer, *Le multilinguisme dans la normalisation de l'internet*.

Mixture Encoding Style. Pour cette raison, les moteurs de recherche les plus courants ne peuvent pas effectuer leur travail correctement à partir de mots birmans.

Les problèmes de segmentation des mots composant les requêtes utilisateur

Chaque langue dispose de ses propres normes en termes de segmentation des mots. Par conséquent, une attention toute particulière doit être accordée à la méthode de segmentation utilisée dans le processus d'indexation. Les moteurs de recherche ont encore du chemin à parcourir pour aboutir à une segmentation appropriée.

Le birman/myanmar

La segmentation des mots est encore plus complexe dans les langues asiatiques, que ce soit en chinois ou en birman, puisque les mots ne sont pas séparés par des espaces. Foo et Li (2004) ont conduit des expériences pour évaluer l'impact de la segmentation des mots chinois sur l'efficacité de la recherche d'information (RI). La précision passait de 0.34 à 0.47 (sur une échelle de 0 à 1) selon la méthode de segmentation utilisée.

De la même manière, nous comparons ici des requêtes en langue anglaise avec des requêtes dans une langue asiatique (myanmar). Dans le cas d'une recherche en langue anglaise, il est de pratique courante pour les moteurs de recherche de retourner également comme résultat à l'utilisateur les pages contenant un sous-ensemble des mots composant la requête initiale. Par exemple, lorsqu'un utilisateur entre les mots « *chocolate ice-cream* », le moteur de recherche affiche non seulement les pages web qui contiennent l'expression complète mot pour mot, mais également les pages contenant séparément les mots « *chocolate* » et « *ice-cream* ». Cela est rendu possible par la tokenisation des données composant la requête. En revanche, lors d'une requête avec des mots birmans dans un moteur de recherche courant, ce dernier se comporte de façon équivalent à une « recherche d'expression complète » (*phrase search*) en langue anglaise. Autrement dit, quand les doubles quotes (" ") entourent la requête, indiquant au moteur de recherche de considérer les mots dans cet ordre exact sans aucun changement. Par exemple, un document A contient un mot composé birman XYZ. Et un second document B contient tous les composants du mot XYZ

sous une forme non consécutive, tel «X....., Y...; Z...»). Si un utilisateur formule sa requête par les termes XYZ, le moteur de recherche affichera bien l'adresse du document A, mais pas celle du document B parce que la segmentation des mots de la requête n'est pas effectuée par le moteur de recherche. C'est pour cette raison qu'un traitement particulier doit être mis en œuvre pour le birman dans les moteurs de recherche. Le myanmar exige une segmentation des mots à la fois dans la phase d'indexation et dans la phase de formulation de la requête.

Pour les langues majeures comme le chinois, ou le japonais, les moteurs de recherche mettent en place des algorithmes de tokenisation qui réalisent les coupures de mots de manière correcte. Dans les exemples suivants, les auteurs souhaitent comparer les traitements effectués par les moteurs de recherche dans le cas d'une langue majoritaire et d'une langue minoritaire. Nous utilisons le japonais d'un côté, et le birman/myanmar de l'autre.

Figure 2: Le fonctionnement d'un moteur de recherche dans le cas d'une langue asiatique majoritaire

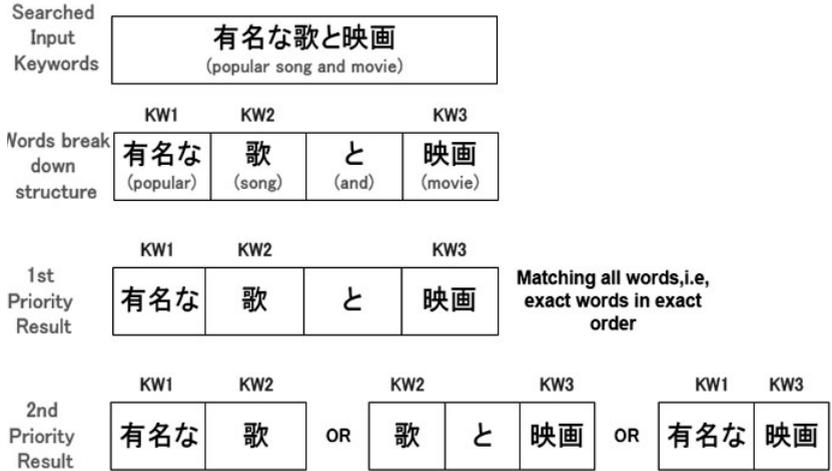


Figure 3: Le fonctionnement d'un moteur de recherche dans le cas d'une langue asiatique minoritaire (myanmar)

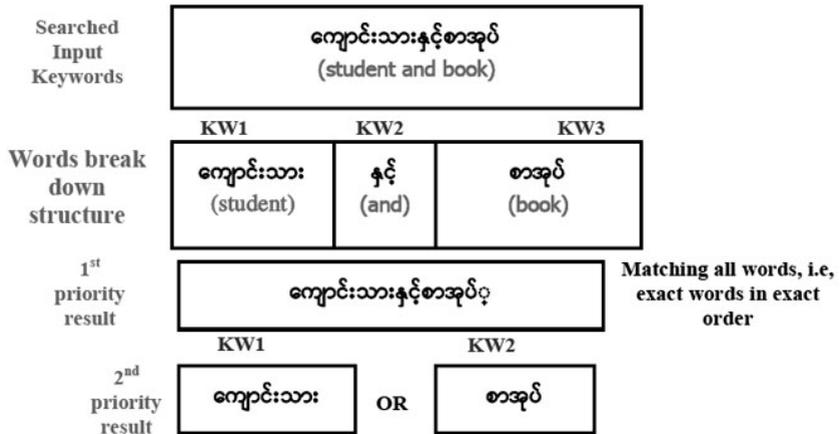
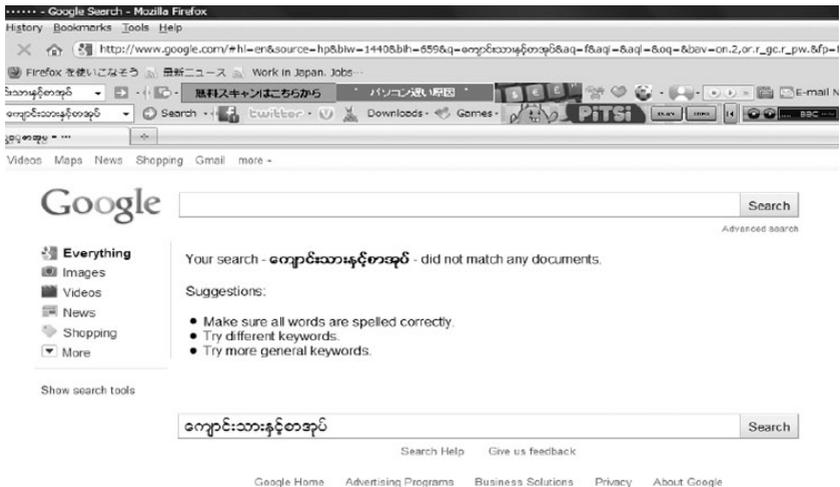


Figure 4 ကျောင်းသားနှင့်စာအုပ်



Comme l'illustre la figure 2, pour les langues majoritaires, les moteurs de recherche restituent la phrase exacte en première priorité. Ensuite, sont extraites les pages web qui incluent les mots segmentés. Les moteurs de recherche fonctionnent ainsi correctement pour la majorité des langues.

La figure 3 illustre l'opération suivie par un moteur de recherche lorsqu'une requête est formulée dans une langue minoritaire. Lorsque les données entrées par l'utilisateur sont « ကျောင်းသားနှင့်စာအုပ် », le résultat devrait afficher les pages qui incluent « ကျောင်းသား » ou « စာအုပ် ». Au lieu de cela, le moteur a effectué sa recherche sur les mots tels que présentés dans la requête de l'utilisateur. Si aucune page n'inclut les mêmes mots alignés exactement dans cet ordre, le résultat affiché indiquera « *Aucun document ne correspond aux termes de recherche spécifiés* 'ကျောင်းသားနှင့်စာအုပ်' » comme l'illustre la figure 4. Il est évident que la plupart des moteurs de recherche ne procèdent pas à des tâches de traitement automatique du langage naturel pour les langues minoritaires.

La langue thaï

Une étude récente menée sur les moteurs de recherche web à propos des requêtes formulées en thaï montre que la segmentation pose toujours problème. D'après l'étude de Virach Sornlertlamvanich *et.al.* intitulée *Evaluation of Web Search Engine with Thai Queries*¹⁰, lorsque le mot « ขำ » est soumis en requête à plusieurs moteurs de recherche, la plupart d'entre eux retournent des résultats qui ne contiennent pas ce mot. Ce qui tend à prouver que la majorité des moteurs de recherche ne gèrent pas correctement la segmentation des mots en thaï.

Les mots composés

Les mots composés présentent une autre difficulté importante pour la plupart des principaux moteurs de recherche lors de la phase indexation des pages web. Certaines langues asiatiques utilisent abondamment les mots composés. Nous prenons ci-dessous l'exemple de mots composés dans la langue birmane, ainsi que dans la langue indienne kannada et en thaï.

Le birman/myanmar

Dans la langue birmane, deux mots simples ခဲ (diriger) and ဝံ (bague, bâton, etc.) se combinent pour former le mot composé ခဲဝံ (crayon).

¹⁰ Virach Sornlertlamvanich, Shisanu Tongchim and Hitoshi Isahara "Evaluation of Web Search Engines with Thai Queries" *Proceedings of Workshop on NTCIR-6 and EVIA-1, NII*, National Center of Sciences, Tokyo, Japan, May 15-18, 2007. <http://research.nii.ac.jp/ntcir/ntcir-ws6/OnlineProceedings/EVIA/15.pdf>

De la même manière, ကျန်းမာ (en bonne santé) se combine avec ချော်ရွှင် (heureux) pour former ကျန်းမာချော်ရွှင် (en bonne santé et heureux). Bien que les mots composés ne soient pas spécifiques à la langue birmane, et soient même largement présents dans toutes les langues, ils diffèrent beaucoup d'une langue à l'autre.

Les Langues Indiennes

De même pour le kannada, langue dravidienne, deux mots simples « ಸು » (bon), et « ಆಲೋಚನೆ » (penser) s'articulent pour former le mot composé « ಸುಲೋಚನೆ » (intellectuel).

Le thaï

Au regard de l'étude réalisée par Virach Sornlertlamvanich *et. al.* (*op. cit.*), certaines requêtes en thaï sont considérées comme des unités indivisibles, bien que chacune d'entre elle puisse être appréhendée comme un ensemble de mots. Un exemple avec la requête « กรมอุตุนิยมวิทยา » trouvée dans le registre des requêtes et qui signifie « Département de Météorologie Thaïlandais ». Cette expression composée peut être considérée en deux mots « กรม » (département) et « อุตุนิยมวิทยา » (météorologie). Puisque ce mot représente une entité unique, il doit cependant être interprété comme une unité indivisible. Certaines requêtes peuvent également ressembler à ce mot, mais la confusion est dans ce cas liée à une mauvaise orthographe. Ainsi, au moins trois requêtes trouvées ans le même journal de requêtes peuvent être considérées comme se référant à ce mot : « กรมอุต », « กรมอุตุนิยม » and « กรมอุตวิทยา ». L'utilisation de ces mots clés renvoie généralement à des sites internet qui utilisent une forme mal orthographiée du mot « กรมอุตุนิยมวิทยา » plutôt qu'au site internet du Département thaïlandais de Météorologie.

Le prétraitement des différents systèmes d'écriture pour un mot donné

Dans toutes les langues, il peut exister différents systèmes d'écriture pour écrire le même mot. Nous prenons ici l'exemple de la langue birmane.

Le birman/myanmar

Le système d'écriture de la langue birmane a été fortement influencé par le pali et le sanskrit. Autrefois, les mots étaient écrits sur des pierres dans des formes simplifiées en raison du peu d'espace disponible. Plus tard, certaines de ces formes simplifiées ont évolué vers des formes étendues, mais peuvent toutefois toujours s'écrire dans leur version simplifiée. Toute personne peut écrire ces mots comme bon lui semble.

Plus encore, certains mots de la langue birmane peuvent s'écrire sous différentes formes sans que soit omis un seul caractère dans aucune de ces formes, même s'ils font partie de groupes de consonnes différents. Les deux formes ont exactement le même sens, la même prononciation mais des représentations différentes. Un exemple avec «ထွင်း» dans sa forme simplifiée. Elle peut être développée sous la forme «ထွင်း» (riz). De la même manière, le mot «သို့» est parfois une abréviation de «သမီး» (fille). Ces abréviations ne sont pas des termes natifs, mais sont le résultat d'une volonté de trouver des formes brèves.

Par conséquent, si de tels mots sont employés dans un moteur de recherche, leur forme étendue doit être traitée comme une phrase.

Les problèmes de racinisation

Pour rendre plus efficaces l'extraction des données, l'une des pratiques mises en œuvre par les moteurs de recherche est la racinisation. Ici, des variantes morphologiques diffèrent d'une langue à une autre.

La famille des langues d'Indonésie

Le malais utilisé en Indonésie est riche morphologiquement. Il y a environ 35 affixes standards (préfixes, suffixes, circumfixes et quelques infixes hérités du javanais)¹¹. En indonésien, les affixes peuvent être attachés quasiment à tous les mots et ils peuvent être combinés de manière itérative. L'usage courant d'affixes semble avoir créé une tendance chez les locuteurs indonésiens à inventer de nouveaux affixes et règles d'affixation¹².

¹¹ Kridalaksana, Harimurti., *Pembentukan Kata Dalam Bahasa Indonesia*. PT. Gramedia, Jakarta 1989.

¹² Tim Penyusun Kamus, *Kamus Besar Bahasa Indonesia*. 2ed. Balai Pustaka, 1999.

En malais, il existe quatre types d'affixes, qui sont les suffixes, les préfixes, les paires préfixes-suffixes et les infixes. Contrairement à un lemmatiseur (algorithme ou application chargée de la racinisation) en anglais qui pourra se contenter de retirer seulement les suffixes pour obtenir les racines, un lemmatiseur qui se veut efficace et puissant en malais doit non seulement être capable d'enlever les suffixes, mais également les préfixes, les paires préfixes-suffixes et les infixes¹³. Si l'ensemble de ces affixes n'est pas traité, les racines ou lemmes, ne pourront pas être exploités pleinement pour une indexation efficace des documents en malais.

Le birman/myanmar

Dans le cas de mots de la langue birmane, la racinisation porte sur la suppression des suffixes inflectionnels, des suffixes dérivationnels, des préfixes inflectionnels et des préfixes dérivationnels. Si l'on s'en réfère à la grammaire officielle du birman, il semble qu'il y ait quatre-vingt onze affixes différents pour les quatre classes principales de mots: les verbes, les noms, les adjectifs et les adverbes. Un lemmatiseur dans la langue birmane doit opérer la suppression directe de ces affixes pour obtenir la racine correcte. La recherche menée à ce sujet par Ko Oo et Yoshiki Mikami donne de plus amples détails sur les lemmatiseurs en langue birmane¹⁴.

L'algorithme de racinisation est différent pour chaque langue. Le moteur de recherche doit tenir compte de ces différences pour traiter chaque langue singulièrement.

CONCLUSION

Il serait beaucoup plus efficace pour les moteurs de recherche de prendre en considération les propriétés singulières de chaque langue et d'étudier les comportements des utilisateurs en situation de formulation de requêtes.

Alors que le Web devient de plus en plus multilingue, et que d'autres langues que l'anglais y gagnent du terrain, le besoin de développer des

13 F. Ahmad, *A Malay Language Document Retrieval System: An Experimental Approach and Analysis*, Universiti Kebangsaan Malaysia, Bangi, 1995.

14 San Ko Oo, Yoshiki Mikami, *Development of Myanmar Language Stemmer*, Master thesis of Management of Information system engineering department, Nagaoka University of Technology, Japan, 2010.

moteurs de recherche adaptés devient de plus en plus évident. Il serait peut-être ambitieux à cette étape de suggérer que les moteurs de recherche s'équipent d'une boîte à outil linguistique complète, capable de gérer toutes les langues, mais tendre vers cet objectif-là n'est pas une aberration, loin de là. Les développeurs de moteurs de recherche devraient s'appliquer à poser les bases de moteurs véritablement multilingues. Les pages web publiés dans des langues qui ne partagent pas les caractéristiques linguistiques de l'anglais peuvent plus facilement passer à la trappe ou être indexées de manière incorrecte par les moteurs principaux.

De manière générale, nous pouvons affirmer que le traitement et la recherche à partir de textes non anglais pose des difficultés supplémentaires de celles apparaissant dans le traitement et la recherche à partir de textes anglais. Les moteurs de recherche doivent par conséquent être localisés, c'est-à-dire adaptés aux langues locales des utilisateurs.

BIBLIOTHÈQUES NUMÉRIQUES

Comment préserver les cultures et les traces des diverses langues au sein des bibliothèques numériques ? Comment valoriser les traductions et faire cohabiter plusieurs versions d'une même œuvre afin que les internautes, notamment les jeunes, puissent concevoir la diversité et la richesse des productions de la pensée ? Comment participer localement, avec sa langue et sa culture, à la construction d'une immense bibliothèque interconnectée, proposant à tou(te)s l'accès aux œuvres du monde entier ?

Article original en français.



HERVÉ LE CROSNIER est maître de conférences à l'Université de Caen Basse-Normandie, où il enseigne les technologies de l'internet et la culture numérique. Il est actuellement en délégation auprès de l'ISCC (*Institut des Sciences de la Communication du CNRS*). Sa recherche porte sur l'impact de l'internet sur l'organisation sociale et culturelle, et l'extension du domaine des biens communs de la connaissance. Hervé Le Crosnier est un des fondateurs de C&F éditions.

HERVÉ LE CROSNIER

NUMÉRIQUES BIBLIOTHÈQUES

Des documents ont été retrouvés partout où la connaissance et la culture a pu être transcrites sur un support. Des collections de papyrus d'Égypte, des tablettes d'argiles couvertes d'écriture cunéiforme de Mésopotamie, jusqu'à la fameuse Bibliothèque d'Alexandrie dont la perte dans les flammes reste un moment fondateur pour ceux qui souhaitent transmettre les savoirs aux générations futures, comme de bénéficier des connaissances des générations antérieures, le désir de regrouper, organiser et rendre disponible les documents a été une activité essentielle des lettrés. Quand l'Europe a vu naître l'imprimerie par caractères mobiles, qui multiplia rapidement le nombre de documents disponibles et favorisa l'échange des livres, on s'aperçut de la nécessité de créer des bibliothèques de dépôt légal pour conserver et accumuler le savoir ainsi constitué. Les premiers enregistrements sonores, rouleaux puis disques du début du xx^e siècle, furent déposés dans des phonothèques.

Pour les langues orales, l'usage et la conservation vont de pair : elles se transmettent directement par leurs locuteurs. L'entrée des langues uniquement orales dans les bibliothèques est récente, renvoyant à l'existence et la généralisation des moyens d'enregistrement audio et vidéo¹. Ces supports multimédias rencontrent dans ces lieux les cultures de l'écrit, qui sont devenues pérennes en s'inscrivant sur des supports. Ils peuvent alors hériter du savoir-faire organisationnel construit autour du livre.

Conserver, organiser, rendre accessibles tous les documents sont les piliers des bibliothèques, qui ont ainsi pu, avec ces fondamentaux professionnels, accompagner chaque nouveau processus d'enregistrement des connaissances et des émotions.

1 Voir dans cet ouvrage : Tunde Adegbola, *Multimédia, langue des signes, langues écrites et orales*.

Aujourd'hui, les documents connaissent pour l'essentiel un « moment numérique », ce qui souvent fait changer l'approche que nous en avons en termes de pérennité et de transmission. Leur création sur traitement de texte fait disparaître le manuscrit. Les pages web sont souvent considérées comme des services, offrant actualités ou commerce, réactions et commentaires, mais trop souvent dans un flux permanent, contraire à la logique d'accumulation des bibliothèques. D'un autre côté, les performances (concerts, événements locaux ou mondiaux, lectures publiques) et même la vie quotidienne (multiplication et généralisation de la photographie numérique et des caméras vidéo domestiques) sont enregistrées. La gamme des supports s'élargit, permettant de rendre pérenne et transformer en document sonore, vidéo ou multimédias les traces de l'activité culturelle et scientifique.

Les bibliothèques numériques se placent à ce carrefour nouveau entre la multiplication des documents liés à la facilité de production et de diffusion de supports numériques et la tradition des bibliothèques. Nous allons essayer ici de définir les bibliothèques numériques, en les distinguant des autres formes d'accès aux documents, puis évaluer les approches et les besoins en terme de multilinguisme. Enfin, nous élargirons sur les contraintes juridiques, techniques, et les nouvelles pratiques culturelles qui encadrent l'activité des bibliothèques numériques.

BIBLIOTHÈQUES ET ARCHIVES

Traditionnellement, nous distinguons trois types d'organismes :

- les bibliothèques conservent et rendent disponibles des « duplicats », ouvrages existant en de multiples exemplaires et ayant été diffusés publiquement, en général par l'édition, parfois par la reprographie pour les rapports ou les documents scientifiques (thèses, « littérature grise »...). Avant de trouver leur place dans les bibliothèques, une première sélection de contenu a été faite par le circuit éditorial, et les matériaux qui entrent dans la bibliothèque sont ainsi largement homogènes par leur approche de la relation auteurs/lecteurs et par le circuit industriel permettant de transmettre le travail des premiers en direction des seconds.
- Les archives conservent les documents internes à une structure (entreprise, institution...) ou un individu (archives personnelles). En

général, les archives traitent des « unicats », documents n'existant qu'en un seul exemplaire, et les gèrent au travers de regroupements organisés par l'archiviste : dossiers, boîtes d'archives, ou ensembles de documents ficelés... Ces ensembles sont souvent les seuls à être décrits dans le catalogue. C'est pourquoi les archives recèlent toujours des documents encore inconnus, et que les curieux y trouvent des trésors. Par extension, on a parlé d'archiver les médias, considérant que le flux audiovisuel ou sonore était un exemplaire unique, même si les constituants (les films, les séries ou les enregistrements musicaux) étaient des documents multiples. Un tel flux était, avant les performances actuelles des « nuages de serveurs », impossible à conserver dans sa totalité. Les archives pratiquent donc « l'échantillonnage » pour ne garder que des exemples (sélectionnés ou pris de façon aléatoire) qui éclaireront les historiens du futur sur les pratiques sociales de l'époque archivée.

- Les musées eux aussi conservent des pièces uniques, mais les accompagnent d'un dossier de suivi, qui retrace l'histoire connue de l'objet, ses divers propriétaires, les restaurations qui ont été effectuées...

Or, ces distinctions, qui renvoient à des approches différentes des documents, deviennent de plus en plus ténues avec le numérique et l'internet.

La principale caractéristique du document numérique est de pouvoir être dupliqué pour un coût marginal presque nul. Les coûts principaux portent sur la création (le coût du prototype, que ce soit celui de l'œuvre originale, ou de la numérisation d'œuvres existantes), et sur l'infrastructure (depuis les centres de serveurs jusqu'aux terminaux des usagers en passant par le réseau de communication). Pourtant, une grande partie des documents ne sont jamais dupliqués : les pages de blogs, les sites de journaux, les catalogues commerciaux en ligne, les photographies déposées sur des services comme Flickr ou Picasa... L'existence d'un point d'accès unique et centralisé permet aux éditeurs de sites de rentabiliser la production et l'infrastructure, soit par la vente de publicité, soit par des abonnements. Les bibliothèques, de par leur expérience collective et multi-centenaire, répartissent le stockage des documents, afin de les rendre à la fois pérennes et accessible au plus près des lecteurs. Ceci fait que le web dans sa globalité fonctionne plutôt comme une archive, composé principalement d'unicats, malgré un coût de duplication faible. On parle d'ailleurs « d'archivage du web » pour décrire l'activité qui consiste à copier sur des disques de très grande taille des échantillons extraits à

périodicité régulière de sites web. Ou du moins de copier leur apparence à un moment donné, de façon à permettre aux lecteurs du futur d'accéder non seulement au contenu textuel, mais à la « matérialité » de la lecture telle qu'elle existait au moment de la production du document.

Ce n'est donc pas uniquement par un retournement de sens énigmatique que le principal service d'enregistrement de copies du web depuis 1995 s'appelle *Internet Archive*. Il fait d'ailleurs plus que d'archiver le Web lui-même, mais veut rendre accessible sur le Web des « archives » audiovisuelles (collections de films, enregistrements musicaux et livres numérisés). En France, l'archivage du web est réparti entre deux institutions pour bien souligner son caractère contradictoire : la Bibliothèque nationale de France conserve les sites web de l'extension .fr², à dominante « textuelle » ou « éditoriale », et l'Institut national de l'Audiovisuel les sites de « diffusion » en mode flux, qui deviennent de plus en plus nombreux (webradios, télévisions web, sites musicaux...). Tout le flux passant par le web fait document qui a le devoir d'être archivé. Aux États-Unis, la Library of Congress a passé un accord avec Twitter pour « archiver » tout ce qui s'échange sur ce réseau social... C'est-à-dire des messages de 140 caractères et les conditions de leur diffusion (qui émet, à destination de quels « followers », et quelles en sont les rediffusions, soit l'image du graphe social de l'auteur)³.

Cette situation nous permet de mesurer le travail nécessaire pour conserver trace du présent à destination des générations futures. La taille du Web dépasse toujours les prévisions. Les documents déposés sur cet immense réseau mondial n'ont été organisés par personne à l'exception des propres producteurs de l'information. On parle ainsi d'autoritativité⁴ pour désigner ce nouveau phénomène qui conduit les auteurs à devenir les seuls décideurs de la publication, par exemple en cliquant sur le bouton « publier » de leur blog. La phase éditoriale, qui transforme une première fois le document pour en faire un livre, ou un autre support accessible au public, a été remplacée par l'immédiateté de la publication/diffusion. Et les services qui veulent garder la mémoire se trouvent alors dans la

2 Ce qui ne prend pas en compte l'ensemble des sites publiés en France, mais environ un tiers.

3 Olivier Ertzscheid, Twitter, un patrimoine superflu(x)? *Affordance*, 9 mai 2010, <http://affordance.typepad.com/mon-weblog/2010/05/twitter-le-patrimoine-du-superflux-.html>

4 Autoritativité, In : *Dictionnaire des concepts info-documentaires*, Savoir-CDI <http://www.cndp.fr/savoircdi/index.php?id=593>

situation de remplir un tonneau des danaïdes : qui doit « juger » de la « valeur » d'un document et l'enregistrer pour le futur ? Ceci n'a jamais été le choix et la mission des bibliothécaires. Doit-on laisser ce rôle à un algorithme qui va mesurer et évaluer en fonction des usages (nombre de clics, nombre de liens...) au risque de laisser dans l'oubli des œuvres majeures que seul le temps et la patience des lecteurs peut mesurer ? Ceci est pourtant arrivé si souvent auparavant⁵ que nous savons que la popularité n'est pas forcément synonyme de qualité. Enfin, dans cet univers de la performance algorithmique, comment résisteront les documents écrits ou dits dans des langues minoritaires ? Le modèle des médias ne peut pas servir de guide pour les actions des bibliothécaires/archivistes.

LES FONDAMENTAUX DES BIBLIOTHÈQUES

Les « bibliothèques numériques » se distinguent donc des « archives du web ». Il ne s'agit pas de conserver le Web, mais d'en extraire le contenu documentaire, et de l'organiser pour des lectures ultérieures. Une manière de prendre la mesure de la numérisation du monde en la rapportant aux documents eux-mêmes. Construire des bibliothèques numériques, qui vont stocker, organiser et rendre disponibles ce type de documents devient une urgence pour garantir l'expansion et le partage de la connaissance. Quand nous parlons de stockage, il faut aussi entendre « conservation », si possible pour une longue période, et donc avec un rafraîchissement régulier des formats d'encodage des documents (format des vidéos, des images, et même des livres numériques).

Pour comprendre les fonctions des bibliothèques numériques, il importe de revenir aux fondamentaux des bibliothèques. Une bibliothèque est un organisme public destiné à construire des collections de documents, à les décrire (catalogage et indexation), les conserver et en offrir la lecture à destination d'un lectorat spécifique. Tous les termes ont ici leur importance.

Tout d'abord, la bibliothèque n'est pas régie par des impératifs commerciaux. C'est l'ensemble de la société, par le biais de la puissance publique et donc des financements collectifs, qui éprouve le besoin de constituer des lieux pour garantir l'accès à tous et sur le long terme aux produits du savoir des humains. Ceci, comme toute solution collective, a des

⁵ Samuel Beckett, qui obtint plus tard le Prix Nobel de littérature, ne vendit que 150 exemplaires de sa pièce *En attendant Godot* l'année de sa parution.

inconvénients, notamment un « temps de réaction » qui est différent de l'intérêt et l'attention que peuvent faire naître les médias. Mais cela garantit que les documents qui constituent les fonds de ces bibliothèques ont été sélectionnés pour que chaque idée, théorie, point de vue... et langue, soit équitablement représenté, tout en laissant le temps et la critique faire émerger les documents les plus significatifs.

Ensuite, une bibliothèque constitue une collection de document, un ensemble organisé, cohérent, représentant une volonté. Le numérique, et le web plus encore, a tendance à se juger par la taille des collections. On parle de millions d'ouvrages numérisés, de millions de photographies, de la quantité de vidéos déposées chaque minute, etc. Cette course, cette volonté de « faire du chiffre » est-elle efficace ? N'est-ce pas une dérive liée au simple fait que ce soit devenu techniquement possible ? Une collection en revanche poursuit un objectif. Elle cherche à assurer soit la complétude d'un domaine restreint (cas des bibliothèques de recherche), ou la satisfaction d'un public (la diversité nécessaire pour couvrir les besoins au sein d'une bibliothèque d'un quartier populaire). Dans le cas du numérique, on doit distinguer le moment de constitution des collections de celui de l'accès aux documents contenus dans ces collections. Ce dernier peut se faire par le catalogue propre de la bibliothèque numérique, mais surtout par le biais de catalogues collectifs, un rôle souvent assuré par des acteurs extérieurs aux bibliothèques constituant les collections. Les moteurs de recherche peuvent indexer les documents de plusieurs collections. Des protocoles sont mis en place pour favoriser la constitution d'index globaux malgré la spécificité de chaque collection. Ainsi, le protocole OAI-PMH⁶ permet à des moteurs de requêtes externes de créer des index à partir des métadonnées relatives à chaque document d'une collection qui est ouverte à un tel « moissonnage ».

Enfin, les collections des bibliothèques sont en relation avec un public spécifique. Les missions d'une bibliothèque d'Université, notamment envers les étudiants, ne sont pas les mêmes que celles d'un laboratoire spécialisé, ou dans un autre sens des actions de promotion de la lecture et d'alphabétisation menées par les « bibliothèques de rue » ou des quartiers, ou des lieux d'exclusion (prisons, hôpitaux...). Qui dit public spécifique, dit aussi un personnel et des services adaptés à ce public.

⁶ François Nawrocki, « *Le protocole OAI et ses usages en bibliothèque* », Ministère de la Culture, France, février 2005 <http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm>

Car une bibliothèque, c'est avant tout un ensemble de services, depuis l'accueil jusqu'à la mission de référence, pour aider le lecteur à retrouver les documents qu'il n'aurait pas su découvrir par lui-même.

Avec cette logique, une bibliothèque se définit plus par son public, son activité, son positionnement et son projet que par les documents eux-mêmes... qui sont rappelons-le des « duplicats », existant en de multiples exemplaires dans de nombreux endroits de par le monde. Ne devons-nous pas garder cette approche dans l'univers numérique ? Au lieu de nous concentrer sur les documents, le nombre de pages numérisées, la taille du catalogue, ne devons-nous pas revenir à la question de ce lien entre la collection et le public qui est assuré en dehors de toute pression commerciale par le bibliothécaire. En tout cas, n'est-ce pas la seule approche qui permettra non seulement de conserver les documents émanant des langues et cultures minoritaires, mais aussi de proposer aux locuteurs de ces langues les documents venant d'autres univers linguistiques adaptés à leurs besoins ?

MÉTADONNÉES

La bibliothèque traditionnelle n'est pas seulement un ensemble d'ouvrages, une collection adaptée à son lectorat, mais remplit également deux missions : celle d'inventaire et de description et celle d'organisation de la connaissance incluse dans les ouvrages. C'est au travers de la gestion des métadonnées que nous retrouvons ces deux aspects dans les bibliothèques numériques.

Les métadonnées représentent l'ensemble des informations que l'on peut obtenir « à propos d'un document ». Ce sont donc en premier les informations du catalogage descriptif : mention des auteurs, illustrateurs, préfaciers... ; indications sur les conditions de l'édition (date de publication, éventuelles réédition, collections, mention de l'éditeur, date de la numérisation,..) ; collation de l'ensemble (nombre de pages, spécificités de l'ouvrage, notamment s'il contient des marques représentant son histoire, *ex-libris* de ses propriétaires, méthode de numérisation, format de fichier, durée des enregistrements sonores ou audiovisuels...) et enfin des mentions de la filiation de l'ouvrage en regard des autres et éventuels ouvrages appartenant à la même « famille » (mention de l'édition originale s'il s'agit d'une traduction, ouvrages en plusieurs volumes...).

Le catalogage descriptif va donner un contexte matériel et éditorial à un ouvrage, qu'il soit imprimé, sur support (CD, DVD) ou directement en ligne. Cette question du contexte prend encore plus d'importance quand on passe aux documents numériques. Trop souvent la facilité d'utiliser le texte lui-même pour retrouver des documents fait oublier ce besoin de replacer un document donné dans un contexte plus général (date et conditions de publication, type et genre de document...). Ajoutons que le numérique, avec la capacité qui lui est propre de relier des informations permet d'aller encore plus loin que les habitudes des bibliothèques. Ainsi, des liens vers des biographies d'auteurs, des photographies, la reproduction des couvertures de toutes les versions linguistiques d'un même ouvrage, ou une liste de lectures critiques disponibles peuvent augmenter cette mise en perspective d'un ouvrage, le replaçant dans l'ensemble de l'aventure éditoriale.

Les métadonnées sont aussi les informations décrivant ou résumant les connaissances contenues dans l'ouvrage et assurant le regroupement d'ouvrages ou de données traitant des mêmes sujets. En premier lieu la « classification », c'est-à-dire le placement d'un ouvrage donné dans le même champ sémantique que d'autres. De cette façon, on peut balayer un sous-ensemble du savoir pour en mesurer la complexité, pour y découvrir des œuvres ayant un regard particulier à côté de grands classiques du domaine. Toutes les spécialités scientifiques disposent de classifications adaptées à la production de documents de leur secteur de connaissance. À côté de la capacité à regrouper des classifications, qui favorise la découverte par inadvertance ou sérendipité, on trouve la précision des mots-clés ou tags qui signalent qu'un document traite d'un sujet précis, susceptible d'être recherché par un lecteur. En ce domaine, on trouve deux approches, qui peuvent se compléter dans le cas des bibliothèques numériques :

- suivre des règles d'établissement des descripteurs, soit en les choisissant dans une liste fermée, préexistante et partagée entre plusieurs bibliothèques (notion de « liste d'autorité »), soit en respectant des cadres préétablis (par exemple le document parle-t-il d'un lieu particulier, d'une personne, d'un événement, d'une période historique...).
- laisser chacun décider des tags, éventuellement en ajouter à la liste proposée. On parle alors de « folksonomie ». Ce modèle fait de chaque tag/descripteur une information indépendante, apportée par les lecteurs eux-mêmes, ce qui pose des problèmes de validation, disperse la

recherche, mais enrichit considérablement les descriptions par l'action des amateurs anonymes⁷.

Là encore, le numérique et le réseau permet de faire cohabiter toutes ces approches. Les professionnels peuvent faire évoluer leur métier pour susciter et valider l'expérience et les savoirs des lecteurs, s'appuyer sur les tags indépendants pour construire des descripteurs structurés. Ils offrent alors un cadre de travail pour permettre aux amateurs et passionnés d'exercer leur volonté collaborative de partage du savoir.

Au final, les métadonnées sont comme des fiches descriptives qui accompagnent le livre. On peut ensuite constituer un catalogue en regroupant ces fiches et en les rendant interrogeables. On peut aussi faciliter la découverte par une circulation entre des fiches succinctes avant de plonger dans la lecture de l'ouvrage ou document en entier. Enfin, les métadonnées permettent de suivre l'histoire d'un document, tant du point de vue physique (par exemple décrire l'histoire d'une œuvre retrouvée qui a été numérisée, comme les manuscrits de Tombouctou⁸) que du point de vue intellectuel (lien avec d'autres versions linguistiques, avec des versions sonores ou des films, accès aux documents préparatoires, ou même aux archives numérisées des manuscrits).

Fiche séparée du document lui-même, le recueil des métadonnées peut aussi favoriser le multilinguisme et l'intercompréhension. Indépendamment de la langue du document, la fiche descriptive peut offrir une première approche multilingue, et permettre ainsi la découverte d'un ouvrage dans des langues étrangères. C'est particulièrement utile pour faire connaître les recherches scientifiques écrites dans des langues locales. Si la fiche de métadonnées est traduite dans les quelques langues majeures de la communauté scientifique concernée, un document destiné aux usagers locaux (étudiants, jeunes chercheurs, société

7 La fécondité de cette approche a été montrée dans le cas des photographies patrimoniales, quand des communautés d'anonymes, en ajoutant de l'information aux photographies, permettent aux bibliothécaires d'améliorer la précision de l'information. L'expérience de la Library of Congress ayant mis de très nombreuses photographies sur Flickr est significatif, comme celle de PhotosNormandie, mis en place par un collectif d'historiens de la bataille de Normandie :

Patrick Peccatte, PhotosNormandie a cinq ans – un bilan en forme de FAQ, *Culture Visuelle*, 27 janvier 2012 <http://culturevisuelle.org/dejavu/1097>

For the Common Good: The Library of Congress Flickr Pilot Project, 30 octobre 2008 http://www.loc.gov/rr/print/flickr_report_final.pdf

8 Tombouctou Manuscripts Project <http://tombouctoumanuscripts.org/fr/>

civile ou décideurs) peut alors s'inscrire dans le champ global du savoir. Et éventuellement recevoir ultérieurement des traductions, en fonction de l'intérêt qu'il va susciter.

Les métadonnées méritent d'être échangées entre plusieurs systèmes, d'être mobilisées différemment pour remplir des besoins spécifiques. Elles sont à la base du web sémantique. À ce titre, elles s'écrivent avant tout dans un format lisible par les ordinateurs. Le modèle le plus répandu actuellement est le format RDF (*Resource Description Framework* – Cadre de description des ressources)⁹. Ce format est normalisé et extensible, et peut traiter toutes les langues du monde, chaque information étant précédée d'un code de langue en cas de multilinguisme. RDF est un format de métadonnées lisible par les ordinateurs, mais on manque encore actuellement de logiciels qui permettent de saisir facilement les informations de contexte. La souplesse du format d'une part et l'extension du nombre d'informations que l'on a envie de conserver dans la fiche descriptive de l'autre rendent ces systèmes de gestion de métadonnées encore assez peu ergonomiques.

On se trouve alors dans une contradiction entre les résultats obtenus uniquement par des logiciels comme la chaîne de reconnaissance optique de caractères puis l'indexation dans tout le texte par des moteurs de recherche, et ceux qui sont offerts par un véritable travail sur les métadonnées. La première méthode est de type « industriel », avec un processus en chaîne de numérisation. Elle est économiquement plus efficace, et elle donne de bons résultats dès lors que l'on accepte de voir « *le livre découpé en pages* » selon l'expression de Jean-Noël Jeanneney¹⁰. L'usage des métadonnées est au contraire un travail « artisanal », chaque professionnel pouvant ajouter des informations sur la fiche, notamment des traductions de termes, de titres, de résumés... Un travail qui peut être ouvert aux usagers eux-mêmes dans un processus que l'on appelle de *crowdsourcing*.

Les documents sont souvent traduits, réédités, sous-titrés... Les ouvrages imprimés tendent à disperser les diverses versions linguistiques d'un même texte, tout comme les DVD ou CD pour les vidéos et le son. Mais les catalogues informatiques, basés sur des fiches de métadonnées

⁹ *RDF primer*, 2004, <http://www.w3.org/TR/rdf-primer>

¹⁰ Jean-Noël Jeanneney, *Quand Google défie l'Europe: plaidoyer pour un sursaut*, Fayard/Mille et une nuits, 2005.

extensibles peuvent compenser cette dispersion, et proposer au lecteur une liste des diverses versions. C'est un concept que les bibliothèques ont développé depuis la fin des années quatre-vingt-dix sous l'acronyme FRBR (*Functional Requirements for Bibliographic Records*¹¹). Ce mécanisme nous offre un moyen de découvrir les versions linguistiques d'une même œuvre les mieux adaptées à chacun. Il peut être étendu aux documents numériques, aux articles de revues scientifiques, aux pages de blogs, aux vidéos doublées, ou possédant une animation signée¹²... En sens inverse, adopter ce modèle FRBR pour accéder aux documents va renforcer le désir de traduire. Souvent nous entendons parler d'un document sous un titre ou un autre critère lié à une langue précise, en général la langue originale du document, et chacun le recherche dans cette version. Or il existe peut être une version dans la langue première du lecteur... que l'on peut alors découvrir. Un processus de découverte qui renforce également l'intérêt de traduire puisqu'il y aura des lecteurs de la version traduite.

NUMÉRISATION

La numérisation consiste à reproduire sous forme numérique des documents existant dans l'univers analogique. On numérise ainsi des livres, des films, des vidéos, des photographies, des documents sonores... Numériser implique de maîtriser deux aspects : les formats de fichiers numériques et la transformation du document numérisé en document interrogeable.

En numérisant, on va transformer l'original en une photocopie numérique. On obtient alors une image dont la qualité visuelle correspond aux moyens techniques dont on dispose au moment de la numérisation. Or ceux-ci évoluent très vite. Les images numérisées il y a seulement vingt ans sont loin d'avoir la définition et la qualité de celles d'aujourd'hui. De surcroît, les fichiers dits « bitmap » dans lesquels on conserve l'information de couleur pour chaque pixel (point d'un codage informatique) sont très gourmands en mémoire et difficiles à transmettre. On a donc mis au point des systèmes et formats de compression qui sont plus maniables,

11 *Modèles FRBR, FRAD et FRSAD*, Bibliothèque nationale de France.

http://www.bnf.fr/fr/professionnels/modelisation_ontologies/a.modele_FRBR.html

Barbara Tillett, *What is FRBR? A conceptual model for the bibliographic universe*, Library of Congress, 2004.

<http://www.loc.gov/cds/downloads/FRBR.PDF>

12 Voir dans cet ouvrage : Annelies Brafort & Patrice Dalle, *Accessibilité du cyberspace : Langues des signes*.

comme jpeg ou png. Le même phénomène existe pour le son (mp3, flac, vorbis...) ou la vidéo (mpeg4, ogg-theora...) Un format est toujours un équilibre entre la qualité et la maniabilité. Une des fonctions des bibliothèques numériques va donc être de conserver l'original numérique dans la plus haute définition possible et de gérer les transformations de format pour utiliser au mieux l'environnement informatique tel qu'il est disponible (qualité du réseau, résolution des écrans, etc.).

Rendre les fichiers interrogeables, trouvables, est plus complexe. Par exemple, il s'agit de reconnaître les caractères, puis les mots, les phrases et mieux encore le sens, dans un texte numérisé¹³. Ou encore de transcrire un flux sonore en texte, de repérer les séquences dans une vidéo, les régions dans une image... On pourra ensuite retrouver les documents à partir de mots (recherche textuelle), ou même en présentant des images et découvrir les images les plus proches¹⁴.

Plus complexe encore sont les opérations consistant à ajouter aux fichiers numérisés, rendus maniables par la compression, trouvables par l'indexation automatique, des informations et métadonnées de contexte, qui aident à réorganiser les pages en livre, les collections de photos en véritable catalogue, les vidéos dans leurs diverses versions linguistiques et sous-titrage... Ces différentes actions demandent une intervention humaine, qui plus est coordonnée dans le monde entier pour bénéficier de l'effet de levier du réseau des bibliothèques. Mais ce sont aussi ces opérations qui garantissent que les documents numérisés vont être aussi bien traités que leurs originaux du monde analogique, qu'ils seront conservés sous une forme ordonnée, qu'ils pourront s'adresser à un public et constituer des collections dans les bibliothèques numériques.

C'est aussi dans cette troisième phase, trop souvent négligée, qu'intervient le multilinguisme, utile pour l'indexation (par exemple disposer des noms de villes dans la langue et l'écriture du pays en complément des noms dans d'autres langues) et pour la mise en relation de documents ayant la même source originale (traduction) ou traitant de sujets similaires (classification).

13 Ceci est particulièrement complexe pour certaines langues. Voir dans cet ouvrage : Pann Yu Mon & Madhukara Phatak, *Moteurs de recherche et langues asiatiques*.

14 Voir par exemple <http://tineye.com>

LIVRES NUMÉRIQUES

À côté des pages web et de leur archivage, ou des documents qui sont numérisés pour pouvoir être diffusés sur le réseau, on voit apparaître maintenant les livres numériques. Ces nouveaux objets documentaires héritent des technologies du web la possibilité d'être lus sur des machines, soit des ordinateurs, mais le plus souvent des « liseuses », terme reconnu en français pour des appareils électroniques spécifiques (Nook, Kobo, Kindle) ou pour des tablettes (iPad, Kindle Fire, Samsung...) ou encore pour des ordiphones qui proposent les deux fonctions de téléphone et d'ordinateur de poche (iPhone, Android...). Mais il s'agit avant tout de « livres », ce qui leur donne deux qualités : ils sont mobiles, faciles à transporter et à utiliser en dehors de toute connexion au réseau, et, comme les livres imprimés, l'ensemble de leur contenu « réside entre deux couvertures »¹⁵. Dans un livre numérique, les techniques viennent de l'expérience de l'internet, mais le concept, l'organisation du contenu et la réflexion sur la mise en forme viennent du monde de l'édition.

Les livres numériques sont souvent conçus aujourd'hui comme des compléments, ou des versions spécifiques de livres imprimés. Le présent livre ne fait pas exception, qui est disponible sous les deux formats. Cependant on verra apparaître de plus en plus des livres qui auront les formes d'organisation et la facilité d'échange et de citation des livres, mais ne connaîtront qu'une version numérique. En particulier, des travaux à diffusion confidentielle, des essais, des compilations de textes, des documents évoluant très vite, comme des guides ou des tutoriels, et enfin les manuels scolaires.

Les bibliothèques vont être confrontées à la conservation et mise à disposition du public de ce type de document. Ceci va étendre leurs missions et responsabilités, notamment en développant une logique de prêt à distance. On voit à cet usage naître des livres « chronodégradables », qui ne peuvent plus être lus après une certaine durée de prêt. Les bibliothèques peuvent aussi proposer des liseuses à distance (en général au sein des navigateurs internet).

Mais le grand enjeu pour les langues est de voir publier des livres numériques mono ou multilingues. La facilité de création, la capacité à améliorer le contenu au fil des rééditions, à doter ces livres d'outils facilitant

15 Michel Melot, Nicolas Taffin (ill.), *Livres*, L'œil neuf éditions, 2006, citation p. 27.

la lecture comme des dictionnaires, des glossaires, des notes, facilitent la lecture de livres dans des langues minoritaires. Ces livres numériques peuvent aussi mélanger dans le même support des textes avec des vidéos ou des images. Enfin, certaines liseuses sont dotées de fonctions de lecture synthétique qui permettent l'accès au contenu pour les mal-voyants.

La multiplication des outils de création¹⁶ et la capacité des professionnels du secteur à réellement s'entendre sur les formats des livres numériques sont deux enjeux pour que ces supports se généralisent et puissent aider au développement du livre, notamment dans les pays n'ayant pas d'infrastructure d'imprimerie ou d'accès au papier ni de réseau de librairies. Les bibliothèques ont alors pleinement leur rôle à jouer. Ajoutons que certaines librairies en ligne comme Amazon essaient de faire passer leur prêt promotionnel de livres pour une activité « de bibliothèque », réduisant le travail de celles-ci à une opération commerciale. Ce qui renforce la nécessité pour les véritables bibliothèques qui gèrent des documents numériques de rester des organisations indépendantes du marché des livres et plus largement des documents, mais bien un service pour l'ensemble de la société, qui garantit l'équilibre des idées et les droits des lecteurs.

ENREGISTREMENTS AUDIO ET VIDÉO

Nous disposons de plus en plus d'outils d'enregistrement, ce qui va conduire à la constitution de collections entières de documents sonores ou vidéos, qui vont à leur tour pleinement entrer dans les bibliothèques numériques. C'est-à-dire connaître les trois étapes dont nous avons parlé au début de ce texte : la gestion des formats de compression, leur maintenance et mise à jour. À ces opérations s'ajoute la nécessité de conserver les documents dans la meilleure définition possible pour des retraitements ultérieurs, et la nécessité d'inscrire des métadonnées donnant le contexte et exprimant de manière synthétique les idées et connaissances exprimées pour faciliter la sélection et la lecture de survol.

La multiplication de ces outils d'enregistrement constitue une opportunité pour les bibliothèques d'enrichir et de diversifier leurs collections du point de vue linguistique. L'uniformisation de la grammaire de l'écrit est très loin d'avoir son équivalent dans le domaine oral. Le français par

16 Voir par exemple l'outil en ligne Polifile (<http://polifile.fr>) qui permet de créer des livres numériques au format ePub.

exemple n'a pas la même prononciation, les intonations, les inflexions, et finalement la même saveur selon qu'il est prononcé par des habitants de Paris, de Lille, de Marseille, du Québec, du Sénégal ou des Antilles. Au sein même des langues majoritaires, des différences existent, très fortes, que les grammaires et les travaux académiques n'arrivent pas toujours à canaliser, tant ces manières de parler répondent à des histoires propres au sein d'une même communauté linguistique. Ajoutons que les parlers évoluent aussi très vite, notamment sous la pression uniformisante des médias qui construisent les formes reconnues et distinguées de prononciation.

Cette capacité à construire des archives orales est aussi un moyen de transmettre les savoirs enregistrés dans les langues en voie de disparition, ou de reconstruire au mieux, avec les enregistrements d'époque, des collections de parlers des langues disparues ou en très grand danger¹⁷.

Le collectage¹⁸, opération visant à recueillir les expressions populaires, depuis les contes, les voix, les chants et la musique, a longtemps fait partie des missions des bibliothèques et des musées ethnographiques. La multiplication, la démocratisation (par baisse des coûts) et la capacité à construire ces collections et les rendre disponibles va renforcer ce travail. Il s'agit de « conserver les langues », au sens de garder la trace des évolutions historiques et des variations des phénomènes linguistiques qui se déroulent sous nos yeux. Les bibliothèques, qui ont pour mission et pour culture professionnelle de se projeter dans la dimension temporelle des documents, ont un rôle particulier à jouer en ce sens.

ACCÉDER LIBREMENT AUX DOCUMENTS

Les documents numériques possèdent cette qualité rare de pouvoir être aisément dupliqués et transmis le long des réseaux. Pour défendre un mode de gestion économique traditionnel du document, tel qu'il est issu de la période où les coûts de reproduction étaient élevés, certains estiment qu'il faut réduire cette capacité en installant à l'intérieur des documents numériques des verrous. Il faut alors une clé spécifique pour

17 Voir dans cet ouvrage : Tjeerd de Graaf, *Intérêt des archives orales pour les langues menacées* et Maik Gibson, *Préserver l'héritage des langues éteintes ou en danger*.

18 Larry Rother, *Folklorist's Global Jukebox Goes Digital*, *The New York Times*, 30 janvier 2012 <http://www.nytimes.com/2012/01/31/arts/music/the-alan-lomax-collection-from-the-american-folklife-center.html>

lire les documents numériques dotés de ces systèmes dits de gestion des droits (DRM, *Document Right Management systems*).

Malheureusement, cette manière de faire ne tient pas compte de la réalité des pratiques sociales. Les lois mondiales sur le droit d'auteur ont depuis toujours envisagé de nombreux cas particuliers, regroupés dans les négociations internationales auprès de l'OMPI¹⁹ sous l'intitulé de « limitations et exceptions ». C'est ainsi que la copie à titre privée, les droits d'usages élargis pour les écoles ou les bibliothèques, mais aussi le droit de citation, de caricature, la libre disposition des discours d'actualité prononcés par des personnages influents... sont inscrits comme autant de droits légitimes qui peuvent être mobilisés par les lecteurs ou par les institutions culturelles et éducatives. Or les systèmes de verrous sont très rigides, et ne peuvent pas prendre en compte le type de pratique ou d'institutions qui souhaite utiliser le document verrouillé. Dès lors, on trouve deux tendances : soit en finir avec les DRM et trouver d'autres modes de financement de la création et de l'édition, à l'image de ce qui s'est passé dans l'industrie de la musique quand les majors ont décidé d'abandonner les DRM en 2006 ; soit modifier profondément les lois sur le droit d'auteur pour en faire un droit dit « naturel », du même type que la propriété sur des biens matériels. Il semble malheureusement que la seconde approche soit aujourd'hui dominante, si l'on en croit les lois et les projets qui sont déposés partout dans le monde sur la propriété intellectuelle.

Quand les bibliothèques participent à la création de documents, comme dans les exemples du collectage ou de la numérisation cités plus haut, elles doivent donc se poser la question des droits qui vont être accordés aux lecteurs. Les licences d'usages, notamment les licences Creative Commons²⁰, sont un bon moyen pour élargir les droits des lecteurs et des institutions publiques. Les négociations avec les éditeurs pour obtenir des versions sans verrous pour l'usage des bibliothèques peuvent aussi être envisagées.

L'exemple des publications scientifiques, trop souvent contenues dans des revues très onéreuses, et ce faisant inaccessible par de nombreuses bibliothèques et universités sur la planète, est significatif. Le mouvement dit *open access*, émanant de chercheurs, propose de mettre à disposition

19 OMPI : Organisation mondiale de la Propriété intellectuelle. Les documents officiels sur les limitations et exceptions sont à <http://www.wipo.int/copyright/fr/limitations/index.html>

20 Voir dans cet ouvrage : Mélanie Dulong de Rosnay, *Traduction et localisation des licences Creative Commons*.

de tous, les versions de leurs articles et travaux²¹. De nombreuses institutions de recherche, comme le CNRS en France, le NIH aux États-Unis... demandent à leurs membres de déposer les articles qu'ils rédigent dans des « archives ouvertes » ou de les publier dans des revues en accès libre.

Cette question des moyens juridiques et économiques permettant d'accéder et de partager l'information contenue dans les documents est centrale pour l'avenir des bibliothèques numériques. Partout les bibliothèques se mobilisent pour faire respecter leur approche collective et de service public. L'IFLA (Fédération internationale des Associations de Bibliothécaires) s'est par exemple engagée dans le mouvement mondial dit A2K : *Access to Knowledge*²².

CONCLUSION

Les bibliothèques numériques, l'archivage du web, la création et conservation de nouveaux documents (livres numériques, enregistrements audio ou vidéo) sont des questions complexes, qui sont pourtant essentielles pour l'avenir des langues dans le monde.

Chaque nouveau support apporte son lot d'opportunités pour mieux comprendre, conserver, diffuser, et parfois même faire revivre les savoirs et les cultures dans les langues mêmes de leur création. Mais des règles techniques ou juridiques peuvent oblitérer cette chance.

Les documents sans métadonnées perdent aussi de leur force en ne proposant pas le contexte historique, culturel, éditorial de leur création. La numérisation qui découpe la connaissance en morceaux ne tient pas compte de ce qu'un livre en bibliothèque est beaucoup plus que la somme de ses parties.

C'est pourquoi les bibliothèques ne peuvent répondre réellement à leurs missions qu'en étant au service de toute la société, en étant des services publics ouverts à tous et défendant les droits des lecteurs à un accès libre à l'information, en apportant le complément nécessaire d'indexation,

21 Jean-Claude Guédon, *Connaissances, réseaux et citoyenneté : pourquoi le libre accès ?* In : *Libres Savoirs, les biens communs de la connaissance*, C&F éditions, 2010, p. 67-75.

22 Pour une vue d'ensemble de cette question, qui au-delà des bibliothèques touche aussi l'accès aux savoirs inscrits dans des biens, comme les médicaments ou l'alimentation : Gaëlle Krikorian and Amy Kapczynski, *Access to knowledge in the age of intellectual property*, Zone Book (diff. MIT Press), 2009, p. 246.

de classification, de mise en contexte. Il s'agit pour les bibliothèques de montrer comment le partage des savoirs est aussi une condition de l'expansion du marché des biens culturels, que l'accès ouvert, géré par les bibliothèques est le moyen d'habituer les lecteurs à des pratiques culturelles qui nourrissent ensuite la production éditoriale marchande.

Les bibliothèques comme institutions, et les bibliothécaires comme acteurs, ont des responsabilités, des compétences et des projets que les décideurs doivent encourager et écouter pour renforcer la présence de toutes les langues dans l'univers mental. La culture est un outil de construction de la Paix dans le monde, qui nous donne des moyens essentiels pour faire face aux défis et fléaux²³ qui touchent principalement les pays pauvres... ceux dont les langues d'usage sont les moins bien dotées en moyens éditoriaux et médiatiques et en bibliothèques et archives. La capacitation culturelle, qui passe par l'école et les bibliothèques, mérite toute l'attention si on la considère comme une infrastructure commune au développement durable et pacifique.

23 Voir dans cet ouvrage : Adama Samassékou, *Multilinguisme, Objectifs du Millénaire pour le Développement et Cyberspace*

LOCALISATION DES LOGICIELS : OPEN SOURCE ET MULTILINGUISME NUMÉRIQUE

Que peut faire le citoyen ordinaire pour promouvoir sa langue à l'ère numérique ? Nous montrerons dans cet article que les mouvements open source et open content fournissent une plateforme idéale permettant aux individus de contribuer à l'existence de leur langue. Nous verrons également comment ces mouvements changent la donne pour celles et ceux qui souhaitent mettre en avant leur langue, modifier les perceptions et influencer sur l'espace politique.

Article original en anglais.
Traduit par Alexandre Mussely.



DWAYNE BAILEY est directeur de Translate, projet dédié à la localisation des logiciels libres et open source dans les langues d’Afrique du Sud et à l’accompagnement de projets similaires dans le monde. Il est le directeur de recherche d’un large réseau d’experts en localisation en Afrique, le Réseau africain pour la localisation (ANLoc).

DWAYNE BAILEY

NUMÉRIQUE
MULTILINGUISME
OPEN SOURCE ET
DES LOGICIELS :
LOCALISATION

Comment les ressources « libres » créent-elles des opportunités pour le développement et la promotion des langues régionales et comment influencent-elles le paysage linguistique ?

LOGICIELS, CONTENUS ET APPLICATIONS

Mais tout d'abord, que veulent dire les désignations *open source* et *contenu libre* et comment diffèrent-elles d'un *graticiel* (*Freeware*) ?

Le *logiciel libre* et le *code source libre*, ou *open source*¹, sont, deux philosophies différentes concernant la manière dont devraient être développés les logiciels. Le mouvement des logiciels libres (*Free Software*) a précédé celui de l'*open source*. Ses partisans estiment que la modification du fonctionnement interne des logiciels est un droit fondamental dont bénéficient les utilisateurs de logiciels. Ils défendent également l'idée du partage des applications et de leur copie. De toute évidence, cette conception est différente de celle soutenue par les propriétaires commerciaux qui utilisent des termes comme « pirate » pour qualifier la copie ou le partage des programmes informatiques. D'autre part, les partisans de l'*open source* considèrent que l'accès du plus grand nombre d'utilisateurs au fonctionnement interne des applications informatiques conduit à une meilleure qualité logicielle. Ces deux philosophies et leurs revendications différentes peuvent faire l'objet d'un débat mais il n'en demeure pas

1 Le terme anglais *open source* est souvent conservé tel quel en français ; aussi l'utiliserons-nous dans cette traduction ; nous éviterons la traduction adoptée par les Québécois de « Code Source Libre » pouvant ici être source de confusion. Compte tenu du sujet traité ici, nous n'entrons pas dans la différence entre logiciels libres et *open source*, et garderons le terme homogène pour désigner tantôt l'une ou l'autre des philosophie, qui toutes deux permettent l'apport de l'utilisateur dans le code informatique.

moins que l'objet de leur attention, le logiciel, joue un rôle essentiel dans l'économie d'aujourd'hui.

L'open source et le logiciel libre partagent l'idée maîtresse d'un accès au fonctionnement interne des logiciels ainsi qu'au droit de les transformer. Ces points communs sont d'une importance capitale pour les langues minoritaires ; car si vous avez le droit de modifier les logiciels, vous avez également le droit de changer la langue des textes les composant. L'industrie informatique nomme *localisation* cette traduction de la langue des interfaces utilisateurs d'un logiciel en une langue locale ou régionale.

Si nous prenons pour exemples deux logiciels open source, nous pouvons aisément constater l'impact de ce facteur sur les langues régionales. Firefox est un navigateur web open source utilisé par 30 % des internautes dans le monde². La version existant au moment d'écrire cet article est disponible en soixante-huit langues et quatorze autres sont en cours de localisation.

OpenOffice.org et LibreOffice sont deux suites bureautiques open source compatibles avec Microsoft Office. Leur part de marché est difficile à estimer puisque n'importe qui peut distribuer le logiciel et qu'il n'est pas obligatoire de s'enregistrer pour l'utiliser. Néanmoins, des données recueillies par le site internet [Webmasterpro.de](http://www.webmasterpro.de)³ montrent que leur part de marché en 2010 varie entre 0,2 % (Chine) et 22 % (Pologne). Leurs chiffres semblent indiquer une moyenne d'utilisation de 10 % sur les marchés sondés, ce qui représente plusieurs millions d'utilisateurs. En ce qui a trait aux langues, ces deux logiciels sont traduits en quatre-vingt-cinq langues par des communautés de bénévoles.

Firefox, LibreOffice et OpenOffice.org, ont ainsi l'avantage de mettre à disposition du plus grand nombre et à moindre coût des logiciels en langues régionales.

Il est important de souligner qu'en utilisant une approche traditionnelle de localisation, l'industrie informatique traduit dans trente-cinq langues au plus. Cependant, certaines entreprises informatiques comme Microsoft ont répondu à la pression exercée par l'open source et ont créé des programmes tels que le « pack » linguistique (LIP), permettant d'augmenter le nombre de langues traduites pour leurs interfaces logicielles.

² <http://gs.statcounter.com>

³ <http://www.webmasterpro.de/portal/news/2010/02/05/international-openoffice-market-shares.html>

Pour sa part le contenu ouvert, que l'on associe généralement à Wikipédia, est également composé de textes et de contenus sous licence Creative Commons⁴. Wikipédia est une encyclopédie en ligne, disponible en plusieurs langues. La version anglaise comporte plus de trois millions d'articles. D'autres grandes langues européennes ont elles aussi un nombre important d'articles dépassant le million d'unités.

Bien que les deux langues africaines les plus représentées sur Wikipédia, l'afrikaans et le swahili, aient un nombre d'articles relativement restreint, avec 17 000 pour la première et 21 000 pour la seconde, il n'en demeure pas moins que ces documents sont un ensemble important de savoirs régionaux. Ces deux langues dans Wikipédia représentent à elles seules 6,4 millions et 3,8 millions de mots sur internet (selon des données de mai 2010). De surcroît, ces articles constituent une ressource gratuite et libre pour tout enseignant désireux de travailler dans la langue maternelle de ses étudiants

Arrêtons-nous quelques instants sur les applications gratuites. Dans cette catégorie, nous pouvons inclure les applications Google, (par exemple : Gmail, Google maps, et le moteur de recherche de Google), Facebook, ainsi que d'autres applications de type « médias sociaux ». Même si ces applications sont gratuites, elles ne permettent pas l'accès à leur code source. Néanmoins, la plupart d'entre elles combinent des programmes de localisation développés en interne et par des communautés de bénévoles.

Des communautés d'utilisateurs acceptent de traduire bénévolement ces applications principalement parce que cela renforce l'utilité directe du programme pour eux-même. Même si le comportement des producteurs de ces logiciels et services, des entreprises conglomérats et loin d'être démunis financièrement, peut être interprété par certains comme de l'exploitation, les utilisateurs qui pratiquent ce travail le perçoivent comme équitable, puisque les applications sont traduites en un grand nombre de langues. Après le lancement de sa plateforme interne de traduction, Facebook a eu très vite plus de 100 communautés actives de traducteurs.

Les logiciels open source et les applications gratuites sont disponibles dans plus de langues que les programmes développés via une approche traditionnelle : plus de 70 langues pour Firefox et plus de 100 langues

⁴ Voir dans cet ouvrage : Mélanie Dulong de Rosnay, *Traduction et localisation des licences Creative Commons*.

pour Facebook, à comparer aux 35 langues en moyenne pour l'industrie traditionnelle. Qu'est-ce qui rend cet écart possible et quelle est la réponse des fournisseurs de logiciels propriétaires ?

La raison qui explique cette différence est assez simple : quand sont écartés les obstacles empêchant de participer à une traduction, un plus grand nombre de personnes s'engouffre dans la brèche. Dans le cadre d'un logiciel open source, le processus de traduction est simple et ouvert. Il est donc peu probable que soit tenu à l'écart quiconque souhaite traduire.

Comparons ceci avec le processus traditionnel de localisation des logiciels. L'entreprise décide dans quelles langues elle souhaite traduire ses produits selon le marché ou d'autres sources d'influence. Elle passe contrat de localisation auprès de structures spécialisées. La traduction sera achevée au terme d'un processus long et fastidieux en phase avec les cycles de vie du logiciel de plus ou moins trois ans. Une traduction qui ne prend donc pas en compte les contributions des communautés d'utilisateurs. Au contraire, les logiciels open source ont des cycles de mise à jour de six mois et les langues d'interface sont ajoutées dès que la traduction est disponible. Une telle rapidité de réalisation met évidemment les traductions commerciales en position désavantageuse.

Ce désavantage explique en partie le lancement par Microsoft de son *Local Language Program* (L.L.P.). Ce programme permet à Microsoft de localiser ses logiciels en dehors du processus de mise à jour et peut ainsi publier les traductions après le lancement officiel de la nouvelle version du logiciel. Même s'il semble peu probable que Microsoft veuille traduire dans une nouvelle langue un produit peu de temps avant le lancement d'une nouvelle version. Dans les premiers jours du L.L.P., les traductions étaient réalisées en contrat avec des universités locales, mais en Afrique du Sud pour le moins, elles sont retombées, une fois de plus, entre les mains des compagnies commerciales de traduction.

Le *Local Language Program* de Microsoft a été une excellente réponse à la force communautaire de l'open source et a augmenté le nombre de langues dans lesquelles le Pack Office et Windows sont disponibles de par le monde.

Maintenant que nous avons cerné les démarches et impacts respectifs des applications open source, des applications libres et des fournisseurs de logiciels de bureau, il est important de comprendre les motivations

des communautés de volontaires, avant d'analyser celles des entreprises marchandes.

Pourquoi localiser ?

La localisation est une opération coûteuse. Même si le travail est fait par des volontaires, il faut avant tout considérer le coût d'un point de vue temporel. On peut estimer à plusieurs mois de travail le temps de localisation d'une des applications mentionnées précédemment. Quelles peuvent donc être les motivations des bénévoles qui traduisent ou dirigent un groupe de traducteurs ?

La première motivation réside dans le fait que *les utilisateurs veulent disposer de logiciels dans leur propre langue*. C'est la motivation la plus immédiate, puisqu'elle ne concerne pas directement la valeur ou l'efficacité des logiciels. Elle s'explique par le fait que les utilisateurs se sentent plus confortables avec des logiciels traduits dans leur langue première.

L'histoire d'un programmeur d'Afrique du Sud et de langue maternelle afrikaans qui refusait d'utiliser des logiciels en afrikaans est une parfaite illustration de ce concept de confort. Puisqu'il employait des logiciels en anglais depuis toujours, il trouvait plus simple et plus naturel d'utiliser des logiciels informatiques en anglais. Dans ce contexte, l'afrikaans lui semblait une langue étrangère. Pourtant, il utilisait des distributeurs automatiques de billets en afrikaans. Pourquoi ? Parce qu'il a toujours eu recours à des services bancaires en afrikaans et qu'il éprouvait des difficultés avec des termes techniques anglais propres aux distributeurs automatiques, comme *account*, *statement*, *withdrawal* ou *deposit*. Dès lors, l'afrikaans était beaucoup plus facile pour lui. Le confort est bien la motivation première et la raison d'être de la majorité des traductions volontaires de logiciels.

La seconde motivation vient de l'influence positive des logiciels localisés sur la vie des utilisateurs. L'Unesco et différents chercheurs encouragent l'éducation en langue maternelle, qui présente de nombreux bénéfices quantifiables et déterminants pour l'amélioration de l'éducation des enfants, tels que la participation dans le processus général d'apprentissage et, plus particulièrement, de l'apprentissage des concepts fondamentaux⁵.

⁵ Voir dans cet ouvrage : Marcel Diki-Kidiri, *Le cyberspace dans l'éducation en langue maternelle*.

Ces mêmes bénéfices peuvent facilement être transposés aux logiciels. De la même manière que l'éducation en langue régionale permet aux étudiants de mieux mémoriser des concepts, de participer plus activement et d'améliorer leurs notes, nous pouvons penser qu'une informatique localisée rendra les utilisateurs locaux plus actifs et plus impliqués. Un usager qui ne connaît que les positions des différents items du menu s'épanouira bien moins qu'une personne pouvant lire les différents sous-menus des interfaces dans une langue qu'il maîtrise. Dans la province du Cap Occidental en Afrique du Sud, des séances d'entraînement en xhosa, utilisant des logiciels anglais, ont montré une amélioration des résultats des étudiants et de leur participation.

Ces deux formes de motivation sont liées aux utilisateurs et à leur ordinateur, mais il existe d'autres raisons pertinentes à la localisation qui ont plus à voir avec le développement des langues. Tout d'abord, nous pouvons considérer l'essor des compétences linguistiques.

Ce n'est un secret pour personne que la première localisation, quelle que soit la langue, est souvent de piètre qualité. Ceux qui localisent nient ce constat, et plus particulièrement les traducteurs commerciaux que l'on paie pour produire un travail de qualité. La mauvaise qualité de ces traductions n'a rien à voir avec les compétences linguistiques des traducteurs. Elles sont généralement pauvres parce que les traducteurs n'ont que très peu d'expérience dans la traduction de logiciels, parce qu'il n'existe pas, habituellement, de terminologie dans la langue cible dans ce domaine et enfin parce qu'aucune convention générale de style n'est disponible dans les langues régionales.

Quoi qu'il en soit, la localisation permet aux traducteurs d'améliorer leurs compétences en se frottant à la transcription de logiciels informatiques. L'acte de localisation développe la langue. Il en est ainsi lorsque nous créons des termes pour désigner les « logiciel de traitement de texte » ou « tableur ».

Sans convention de style et sans terminologie, il y a peu de chance que les traductions soient cohérentes. La qualité des traductions passe en premier lieu par le développement préalable de ces deux outils. Malheureusement, quiconque n'a pas encore régionalisé et tente de les développer obtiendra de mauvais résultats. Ce qui nous renvoie à notre première hypothèse : la première localisation sera de piètre qualité. Mais développer des

connaissances et des guides stylistiques s'appuyant sur une pratique de la langue conduit à une amélioration progressive des traductions.

L'open source à cet égard est remarquable. Lorsque les cycles de publication sont lents, le risque existe que les gens finissent par accepter des choix stylistiques et terminologiques de mauvaise qualité. Il existe plusieurs exemples en différentes langues, où le traducteur, sans trop y réfléchir, suit littéralement le style anglais dans sa traduction. L'exemple le plus courant est celui des Majuscules dans les Titres. Les mots des titres anglais ont tous une majuscule tandis que dans d'autres langues la majuscule n'est mise que sur le premier mot d'une phrase. Malgré cela, les intitulés dans un logiciel sont souvent traduits selon la règle anglaise. Le code source libre permet de dépasser facilement ce problème, puisque les erreurs peuvent être corrigées au fur et à mesure de l'émergence d'une terminologie et d'une convention de style dans les langues nouvelles.

La localisation favorise le statut des langues et change les mentalités à leur égard. Pour un grand nombre de langues minoritaires, la question de la traduction conduit souvent à des débats virulents sur la pertinence ou non de traduire un logiciel. On retrouve généralement deux orientations dans le camp des opposants à la traduction : « *Nous n'avons pas besoin de traduction, nous devrions utiliser les logiciels en anglais, en français, etc.* » ou « *notre langue ne peut pas être utilisée pour des logiciels informatiques* ». Or, ces deux arguments contre la localisation n'ont aucune cohérence. Si la plupart des utilisateurs d'une langue donnée utilisent des logiciels dans une autre langue, dominante, c'est précisément parce qu'ils parlent déjà ces langues que cela ne leur pose pas de problème. Leurs arguments sont donc l'expression de leur propre expérience et ne sont pas le reflet des besoins de l'ensemble des locuteurs de leur langue.

Le fait qu'une langue ne dispose pas d'un terme pour définir un concept informatique peut être dépassé en créant ces termes manquants. En vérité, la plupart des mots que l'on cherche à traduire dans une langue comme, par exemple, les mots *proxy*, *password* et *autorisation*, existent déjà. La majorité des mots que l'on rattache à l'univers informatique sont plus communs qu'on ne le croit et sont souvent utilisés depuis des siècles.

D'une manière générale, le débat sur la viabilité d'une langue pour interfacer un logiciel s'atténue après le lancement de la première traduction. La facilité d'usage et l'utilité seront déterminées par l'adoption, ou non,

à long terme du logiciel. Mais qu'une langue ne puisse pas être utilisée pour un ordinateur est un argument fallacieux.

Pourquoi un logiciel open source offre-t-il un meilleur environnement pour la promotion et le développement des langues qu'un logiciel propriétaire ?

La localisation requiert un nombre de compétences qui peuvent être apprises via un enseignement, mais ne peuvent s'affûter que dans la pratique. Même si les efforts fournis par des sociétés comme Microsoft et son *Local Language Program* font émerger des logiciels localisés, ils ne sont pas des terrains stables sur lesquels bâtir ses compétences. Quand une entreprise fait appel à des traducteurs locaux, elle espère que la localisation de son logiciel sera de bonne qualité. Mais pour une langue sans expérience de localisation préalable, les compétences et les ressources nécessaires manquent. Un traducteur professionnel peut éventuellement remplacer une personne expérimentée en localisation, mais il ne sera jamais meilleur que ce dernier.

Pourquoi peut-on affirmer que le *code source libre* est un environnement stable ? A-t-il des normes moins exigeantes ? En réalité, il est stable car les erreurs y sont tolérées et vite corrigées. La qualité s'accroît dès lors qu'on accepte les erreurs et que l'on facilite leur correction.

L'univers open source comme celui des traducteurs professionnels s'assurent de la qualité de leur traduction par certains procédés complémentaires de révision. Tout d'abord, la traduction est revue par un second traducteur. Ensuite, elle est évaluée à l'usage par des traducteurs utilisateurs du logiciel en question. Une telle évaluation n'est possible que si le traducteur connaît les idiosyncrasies propres à sa langue lorsqu'il s'engage dans la traduction logicielle. Rappelons notre hypothèse que sans une première localisation dans une langue, personne n'est en mesure d'identifier ces difficultés. Il faut d'abord s'y être frotté. Si l'évaluation est assurée par quelqu'un qui n'est pas familier avec ce type de problème, elle ne sera que superficielle. De même si l'évaluation de la traduction à l'usage du logiciel n'est pas assurée par quelqu'un qui utilise des logiciels localisés ou qui n'est souvent même pas un expert des logiciels traduits, elle ne sera pas fiable.

En ce qui a trait au développement des langues, l'open source dispose de deux avantages. Premièrement, sa communauté dynamique peut réagir facilement aux problèmes de traduction, de style et de terminologie. Deuxièmement, les cycles de publication varient entre six et douze mois, tandis que ceux des logiciels commerciaux prennent près de trois ans. Ceci permet donc l'émergence rapide de nouvelles traductions. À titre d'exemple dans le champ du logiciel propriétaire, la traduction zulu du *Local Language Program* de Windows XP, après avoir été révisée et lancée sur le marché, n'a jamais été mise à jour. La version suivante en zulu n'a été disponible qu'avec Windows Vista. Quel que soit l'environnement, il existera toujours des erreurs de traduction mais l'open source fournit un environnement d'expérimentation où les utilisateurs peuvent réagir immédiatement, permettant la modification rapide des erreurs. Ce qui évite également que ces erreurs grossières ne se répercutent sur d'autres traductions dans la même langue.

L'open source est à l'origine de développement pour les langues au travers de la création de ressources libres. Bon nombre d'entreprises ont développé des terminologies et des conventions de style avant de débiter une traduction dans une langue nouvelle, deux ressources importantes, comme on a pu le voir, pour localiser.

Toutefois, si ces ressources sont facilement créées pour des langues bien établies, il en est autrement pour des langues minoritaires. Pour les premières, les ressources existent déjà sous diverses formes. Il s'agit donc de les rassembler pour le travail de localisation. Pour ce faire, on peut utiliser les règles grammaticales et orthographiques existantes, ainsi que d'éventuels dictionnaires bilingues. Par contre, la situation est tout à fait différente pour des langues locales ou régionales. On peut toujours créer une convention de style, même s'il n'existe pas de convention grammaticale ou si aucun travail de localisation n'a été mené auparavant. Mais il est nécessaire d'avoir à l'esprit que si c'est un bon point de départ, cela s'inscrit surtout dans un processus de développement et ne doit pas être envisagé comme une fin en soi. Même chose pour la terminologie. Rassembler une équipe de linguistes pour identifier un corpus de six mille termes, sans pour autant comprendre leur utilisation dans un logiciel, restera toujours problématique.

Reste qu'après avoir créé ces ressources, leur accessibilité est le principal problème à résoudre pour une langue émergente. Pour prendre un

exemple propre à l'Afrique du Sud, Microsoft a créé 6000 termes informatiques dans différentes langues de ce pays. Après avoir été évaluée et approuvée par le Comité linguistique pan-sud-africain (PanSALB), la liste de mots n'a pas été mise à la disposition du public et des traducteurs. Cette erreur est à mettre sur le compte du comité, puisqu'il lui aurait été facile de demander le libre accès de la liste, du moment où la révision s'est faite avec des fonds publics. À mettre au crédit de Microsoft, ces listes de terminologies (ainsi que les conventions de style) ont finalement été mises à la disposition du public, mais seulement après plusieurs années.

À l'inverse, l'open source, de par sa nature même, doit créer des ressources libres. Sous une licence de libre diffusion, les guides stylistiques et les listes terminologiques créés pour une traduction d'un logiciel open source peuvent être utilisés et améliorés par d'autres utilisateurs.

En traduisant des logiciels open source, les traducteurs améliorent leurs compétences, en créent de nouvelles et développent de nouveaux marchés. Dans un environnement où les langues dominantes sont acceptées, il n'y a pas de besoin de compétences en traduction. En revanche, dans un environnement où l'on utilise des logiciels localisés, se crée une demande grandissante pour la localisation. Il est donc possible en localisant des logiciels open source de stimuler la demande des utilisateurs vers des logiciels traduits dans leur langue maternelle, ce qui entraîne une demande de la part des producteurs de logiciel et, par conséquent, crée du travail pour les traducteurs.

Le marché se développe et l'open source permet aux traducteurs professionnels d'améliorer leurs compétences. Ils traduisent activement sans encourir le risque économique d'une mauvaise traduction locale. De plus ils travaillent avec des logiciels de localisation, suivent des processus formels en ce domaine et reçoivent des commentaires d'ingénieurs spécialisés dans la localisation, Des compétences qui peuvent être utilisées, ultérieurement dans le secteur du logiciel propriétaire.

Dans un environnement universitaire ou dans celui d'une formation, l'open source s'avère un outil précieux pour enseigner les principes de localisation à des étudiants: il autorise la confrontation à une véritable problématique, à l'opposé de simulations souvent improductives. Et les étudiants sont très conscients que leurs travaux seront réutilisés par d'autres personnes et pour de vrais logiciels.

L'open source et les contenus libres peuvent donner naissance à une communauté linguistique bien plus efficace que des personnes isolées ou des sociétés privées. L'open source est avant tout un mouvement fondé sur le volontariat. Ce fait est important pour les langues régionales, car si les personnes concernées prennent conscience que leurs efforts et le temps consacré ont un impact sur leur propre langue, celle-ci ne risquera pas de disparaître.

Si l'on paie cinq personnes pour traduire chaque jour un article de Wikipédia durant une année, mille articles seront traduits *in fine*. Si dans le même temps cinquante bénévoles traduisent un article par mois, six cent articles seront traduits. La vitesse de traduction n'est pas aussi intéressante dans le deuxième exemple que dans le premier, mais la pérennité de l'initiative est mieux assurée. Chacun des cinquante traducteurs est susceptible de trouver de nouveaux traducteurs alors qu'il est peu probable que les cinq en recrutent d'autres. De plus, une fois le contrat achevé, les cinq traducteurs s'évaporent dans la nature, tandis que la possibilité que les bénévoles continuent est importante. Dans cette perspective, une approche combinée de ces deux groupes aurait un énorme potentiel, en ce sens qu'elle permettrait aux bénévoles de s'appuyer sur le travail des professionnels. Après tout, il est beaucoup plus intéressant d'intégrer un grand projet de traduction de près de mille articles qu'un corpus d'une vingtaine.

De la même manière, un groupe important de traducteurs de logiciels dans une langue locale peut localiser plus de logiciels qu'un ou deux traducteurs isolés.

ANLOC : AFRICAN NETWORK FOR LOCALISATION, RÉSEAU AFRICAIN POUR LA LOCALISATION

L'African Network for Localisation (ANLoc) a été en mesure d'appliquer ces principes de régionalisation propres à l'open source. Au sein même du réseau, l'open source fournit une extraordinaire plateforme pour nos programmes d'action « localiser les logiciels » et « formation ». Au final, notre projet a généré un nombre important de traductions de logiciels en langues régionales. Firefox a été traduit en sotho du nord, en luganda, en akan et en songhay, ainsi que le logiciel de dessin pour enfants TuxPaint, celui de traitement de texte AbiWord, et le lecteur multimédia VLC.

Pendant les sessions de formation de l'ANLoc, nous utilisons le logiciel de messagerie instantanée Pidgin et l'outil de traduction assistée par ordinateur Virtaal pour enseigner les principes de la localisation. Ces deux applications sont des logiciels open source. L'utilisation de Pidgin nous a permis de montrer facilement aux participants les résultats de la traduction. Un des participants a même traduit en entier le logiciel Pidgin en swahili, devenant ainsi un meilleur traducteur dans sa langue locale tout en stimulant la demande pour plus de localisation.



Capture d'écran d'une fenêtre de dialogue des préférences de firefox en swahili et de l'onglet de téléchargement de Firefox en afrikaans. Ces deux langues bénéficient du soutien des initiatives de l'ANLoc.

Après avoir souligné l'importance d'une localisation dans le cas de Pidgin, nous avons reçu les commentaires d'un traducteur ayant localisé des produits de Microsoft et qui venait seulement de comprendre les enjeux liés. Bien que l'open source ne soit pas indispensable pour en démontrer les points clés, il est intéressant de noter que si aucun des participants à la formation n'avait participé au développement informatique de Pidgin, tous purent néanmoins aisément modifier son code source pour ajouter les nouvelles traductions. Les participants à la formation ont pu se confronter à ces problématiques en mettant les mains à un logiciel dont ils savaient qu'il serait utilisé par de vraies personnes.

Dans le cadre du projet *localise software programme* de l'ANLoc, des aides financières ont été allouées à plusieurs équipes de traduction réparties sur tout le continent africain, en charge de traduire des logiciels open source.

Ce programme avait ceci de singulier que les équipes devaient respecter un planning strict de travail. Planning mis en œuvre par des traducteurs expérimentés, et qui permit aux équipes de se concentrer sur leur tâche sans avoir à décider quel logiciel traduire. En outre, une équipe d'experts techniques a accompagné les traducteurs, prenant en charge les aspects techniques d'implémentation qui auraient pu gêner des débutants.

Tout le monde peut régionaliser et c'est ce que démontre l'expérience de l'ANLoc. La qualité des localisations doit bien évidemment être évaluée par les communautés linguistiques concernées, mais nous montrons qu'à l'aide d'experts techniques (capables de rendre invisible la complexité logicielle et d'aider à résoudre des questions pointues de traduction) et de plateformes web ergonomiques de localisation (sans pré-installation et faciles à maîtriser), de petites équipes peuvent avoir un impact majeur sur leur langue minoritaire.

Les plateformes web de traduction distribuée sont le futur de l'open source et des communautés de traducteurs. Au sein de l'ANLoc, nous poursuivons le développement d'une telle plateforme, Pootle, créée par le site internet Translate.org.za. Cette plateforme est d'une importance fondamentale pour les traducteurs bénévoles dans la mesure où elle assure un plus grand accès à ce travail de traduction.

Elle garantit également le partage des ressources, lequel permet la localisation et, à long terme, s'avère très bénéfique pour les langues minoritaires. Pour ces dernières, il est important que listes de terminologies et expertises de traductions soient partagées. Pour les langues dominantes, ce type de partage est moins important. En revanche, pour une langue comme le xhosa, si on crée une liste de termes propres aux technologies de l'information et de la communication, on doit pouvoir la partager et s'assurer qu'elle soit connue et utilisée par les différents traducteurs.

Qu'est-ce qui incite les entreprises à localiser? Elles le font pour augmenter leur part de marché, pour y acquérir un avantage concurrentiel ou pour respecter les directives légales. Le principe est simple: si une entreprise peut vendre plus de produits en les localisant, si elle peut acquérir de l'avance sur son marché ou si une politique publique lui demande de le faire, alors elle localisera ses produits.

Même si les entreprises se facilitent la tâche en disant qu'elles localisent pour répondre aux besoins des communautés, leurs véritables motivations

sont dans le triptyque dépeint ci-dessus. La motivation la plus singulière est celle des opportunités et des freins à un marché. Avec une politique linguistique en Afrique du Sud qui s'applique à onze langues mais aucune politique de traduction de logiciels, il n'y a pas besoin de traduction. Et avec un marché qui n'a pas besoin de logiciels localisés et un faible pouvoir d'achat, il n'y a rien qui incite les entreprises à tailler dans leurs profits pour localiser. Notamment si cette dernière opération n'apporte aucun avantage marché.

Pourtant, quand OpenOffice.org a été traduit en Afrique du Sud, cela a entraîné un désavantage pour ses concurrents sur le marché des suites bureautiques. Afin de reconquérir le terrain perdu, Microsoft a traduit ses logiciels. Si nous faisons abstraction du nom des entreprises, nous voyons un principe général qui peut servir les intérêts des activistes des langues minoritaires. Quand il y a une opportunité pour changer les conditions du marché, l'open source peut être utilisé pour y parvenir.

L'open source permet aux langues minoritaires de former leur propre environnement linguistique. Les programmes de traduction de l'industrie du logiciel étaient fondés sur la valeur marchande avant qu'ils ne changent sous la pression de l'open source et ne prennent en compte d'autres contraintes telles que les politiques linguistiques régionales, le nombre de locuteurs, etc. Les localisations de logiciels open source ont pour effet d'annuler l'argument de l'absence de produits localisés.

Avec leurs politiques linguistiques respectives, des pays comme l'Islande et le Canada se sont assurés que les produits de Microsoft seraient d'un côté traduits en islandais, et de l'autre seraient publiés en même temps en anglais et en français. Alors que la plupart des langues minoritaires n'ont pas d'influence au niveau politique, l'open source leur permet d'en obtenir et de transformer les règles du marché. Quand il n'existe aucun logiciel traduit en une langue minoritaire et qu'il n'y a pas de marché pour ce produit, il est facile d'affirmer qu'aucun logiciel ne devrait être traduit. Or, du moment où les traductions des logiciels open source sont disponibles, elles deviennent aisément des facteurs de développement local pour les entreprises marchandes souhaitant contrecarrer leur handicap sur le même marché.

Si nous regardons le programme LLP de Microsoft, nous pouvons constater que dans la majorité des cas, une nouvelle langue y est introduite

lorsqu'elle a été précédée par une initiative et par une traduction d'application open source. Cela est vrai pour des pays comme l'Afrique du Sud, le Népal, la Tanzanie et d'autres. Même si Microsoft s'est doté du LLP pour répondre spécifiquement à cette pression extérieure, il ne l'utilise effectivement parce que les traductions des logiciels open source l'ont poussé à réagir.

En Afrique du Sud, le groupe de langue afrikaans a demandé à plusieurs reprises des localisations à Microsoft, notamment par l'envoi de lettres officielles et par une rencontre avec le père de Bill Gates. Et, même si la constitution de 1994, qui reconnaît onze langues officielles, a réaffirmé ce besoin, plus d'une décennie fut nécessaire pour que le LLP permette à Microsoft de répondre présent. Toutefois, sur la même période Microsoft a lancé cinq versions du système d'exploitation Windows, qui n'ont à ce jour pas été traduites dans la totalité des langues officielles d'Afrique du Sud.

Remarquons que, trois semaines après le lancement d'OpenOffice.org dans trois des langues officielles d'Afrique du Sud en 2004, Microsoft annonça via un communiqué de presse, qu'il traduirait ses produits dans les langues concernées dans un délai de six mois. La promesse sera tenue au bout de deux ans. On pourrait évidemment les accuser de pratique publicitaire mensongère, mais ce qui nous importe c'est qu'ils ont fini par publier les versions locales. Et cela grâce à l'open source.

La conclusion que nous pouvons tirer de cette situation, n'est pas que Microsoft ne prend pas en compte la diversité linguistique, mais plutôt que l'open source est un facteur d'incitation à la localisation pour les entreprises qui craignent d'être défavorisées sur leur marché. Un activiste traduisant un logiciel open source est beaucoup plus efficace qu'une personne sollicitant publiquement par courrier une localisation à des producteurs de logiciels. L'open source est un outil puissant pour les activistes des langues !

LA LOCALISATION DE L'OPEN SOURCE N'EST PAS CONDITIONNÉE PAR UNE LOGIQUE ÉCONOMIQUE MAIS SOCIALE ET CULTURELLE.

Il est intéressant de se rappeler que les intérêts commerciaux ne correspondent pas toujours aux intérêts de la société dans son ensemble. Les entreprises du secteur privé ne se sentent pas concernées par des langues

régionales parlées par une dizaine de milliers d'individus. Tandis qu'avec l'open source, il suffit d'une personne pour faire d'une priorité individuelle un projet collectif. En d'autres termes, l'open source permet le développement et la pérennisation des intérêts non commerciaux.

Nous pouvons aussi inclure parmi ces intérêts : la volonté de promouvoir et de faire avancer la diversité linguistique, de rendre prioritaire la langue régionale dans le système d'éducation ou encore, de revitaliser des langues en voie d'extinction. Les motivations peuvent aussi être politiques, pour de bonnes ou mauvaises raisons, en faveur du développement des langues. En fin de compte, ces motivations sont prises à bras-le-corps par ceux et celles qui se soucient de leur langue, sans pour autant avoir à justifier leurs démarches par une logique économique.

Même si nous n'avons pas traité en détail de la question des contenus libres dans cet article, nous pouvons dire que la plupart des principes de l'open source sont transposables au contenu libre. La différence étant que l'open source fournit les instruments nécessaires à la traduction, alors que le contenu libre stimule la création d'un ensemble autosuffisant de contenu en langue régionale.

Le message est clair. Pour développer des langues régionales à l'ère numérique, il faut localiser les logiciels open source pour créer un environnement linguistique, une demande, une envie et des compétences qui mèneront tous vers de plus en plus de localisation et, en fin de compte, vers un véritable écosystème numérique en langue régionale.

TRADUCTION ET LOCALISATION DES LICENCES CREATIVE COMMONS

Creative Commons est un ensemble de règles juridiques et de licences destinées à fluidifier la circulation des travaux de création. En faciliter l'adoption légale dans les juridictions locales va bien au-delà de la simple traduction d'une langue à l'autre. Elle demande aussi la construction d'une communauté d'experts de par le monde.

Article original en français.



MÉLANIE DULONG DE ROSNAY est chargée de recherche à l'ISCC (*Institut des Sciences de la Communication du CNRS*) et responsable juridique de Creative Commons France au CERSA CNRS Université Paris 2. Elle a co-fondé Communia, l'association internationale pour le domaine public numérique. Docteure en droit, elle a été chercheuse au Berkman Center for Internet & Society (Harvard Law School) et à l'Institute for Information Law (University of Amsterdam).

MÉLANIE DULONG DE ROSNAY

TRADUCTION
ET LOCALISATION
DES LICENCES
CREATIVE
COMMONS

L'extension de la durée de propriété accordée par le droit d'auteur et de son champ d'application réduit considérablement les possibilités d'accès et de réutilisation des œuvres de création, alors que les technologies numériques peuvent (et devraient) être utilisées pour faciliter leur usage au lieu de les rendre plus inaccessibles encore. Le droit d'auteur octroie automatiquement aux auteurs un droit exclusif de contrôler la reproduction, la distribution et la modification de leurs œuvres, laissant peu de droits disponibles pour le public sans autorisation, comme la parodie, certaines utilisations privées ou éducatives.

Cependant, les créateurs peuvent choisir de laisser d'autres personnes copier et réutiliser leurs œuvres librement. En décidant d'être plus généreux, ils obtiendront une visibilité accrue, peut-être soutiendront-ils la participation citoyenne, le remix créatif ou la traduction par des volontaires. Creative Commons (CC), une organisation à but non lucratif, met gratuitement à la disposition du public un jeu de licences ouvertes qui lèvent les barrières à l'accès et à la créativité en facilitant le partage des créations¹. L'auteur qui distribue son œuvre sous une licence CC autorise le public à copier son œuvre à condition que certaines règles soient respectées, comme indiquer la paternité, ne pas effectuer d'utilisation commerciale sans autorisation supplémentaire ou ne diffuser d'éventuelles adaptations que sous une licence identique à celui régissant l'œuvre d'origine.

Les licences CC peuvent être appliquées aux créations qui sont couvertes par le droit d'auteur : textes, billets de blog, articles, livres, images, sites web, œuvres audiovisuelles, photographies, musique, etc. Elles sont utilisées par

1 L'ouvrage *Net.lang* que vous tenez dans les mains est ainsi distribué sous la licence CC-by-sa (attribution) pour favoriser les traductions et les éditions dans tous les pays et pour toutes les langues.

des artistes individuels et des institutions telles que Wikipedia, Al Jazeera pour des séquences audiovisuelles, le MIT (Massachusetts Institute of Technology) pour des ressources éducatives ou Hindawi (maison d'édition adoptant un modèle de publication en Accès Ouvert) pour des revues scientifiques. D'autres instruments sont disponibles pour les données ou les œuvres du domaine public qui ne sont pas couvertes par le droit d'auteur: la Maison Blanche ainsi que les gouvernements des Pays-Bas et du Piémont utilisent une cco (0 pour zéro) pour indiquer qu'ils renoncent à exercer leurs droits sur les données publiques, en vue de faciliter l'accès des citoyens à l'information et l'innovation basée sur la réutilisation de données publiques.

Les licences CC sont accessibles pour le public à travers une interface en ligne². Celle-ci permet aux auteurs d'indiquer les droits qu'ils souhaitent accorder au public en choisissant des éléments optionnels. Ceux qui offrent une licence peuvent (ou non) demander à ce que leur œuvre soit utilisée seulement à des fins non commerciales ou sans adaptation, ils peuvent aussi demander à ce que les adaptations éventuelles de leur œuvre (par exemple les traductions) soient licenciées selon les mêmes conditions. En fonction des réponses à ces questions, l'utilisateur se verra délivrer l'une des six licences à afficher sur son site web ou sur l'exemplaire physique de ses œuvres afin d'indiquer au public quelles libertés sont accordées à l'avance et quels droits restent réservés.

Les six licences combinant les différents éléments optionnels sont les suivantes:

Licence	Logo	Licence	Logo
Attribution (BY)		Pas d'Utilisation Commerciale (BY NC)	
Pas d'Utilisation Commerciale - Pas de Modification (BY NC ND)		Pas de Modification (BY ND)	
Partage à l'Identique (BY SA)		Pas d'Utilisation Commerciale - Partage à l'Identique (BY NC SA)	

² <http://creativecommons.org/choose/?lang=fr>

L'auteur peut indiquer dans les versions numériques que son document est diffusé sous licence Creative Commons de plusieurs manières :

- Le bouton avec le logo CC, contient un lien vers le résumé explicatif de la licence ;
- Du code informatique en format RDF avec des métadonnées à destination des moteurs de recherche ;
- Un résumé explicatif des libertés essentielles et des restrictions optionnelles contenues dans la licence : <http://creativecommons.org/licenses/by-sa/2.0/fr/> ;
- Le code juridique, soit la licence complète : <http://creativecommons.org/licenses/by-sa/2.0/fr/legalcode>.



Le code juridique a été rédigé à l'origine en référence au copyright américain, l'organisation étant installée dans cette juridiction. Il a été par la suite réécrit en référence aux conventions internationales en matière de droit d'auteur. Les licences ont été traduites dans plus de cinquante langues couvrant plus de soixante-dix pays. Ce processus appelé « transposition internationale » va au-delà d'une simple traduction du texte légal. Par exemple, les définitions proviennent des législations de chaque juridiction.

La localisation des licences permet non seulement une traduction linguistique et juridique, mais améliore de surcroît leur accès, leur acceptabilité et leur compréhension par les utilisateurs ou les juges qui auront à les interpréter dans leur juridiction. Le processus d'internationalisation est également l'occasion de constituer des équipes locales d'experts en droit d'auteur et licences ouvertes. En sus de la prise en charge de la traduction et de l'adaptation des licences, les responsables des équipes affiliées à CC travaillent avec les communautés locales d'utilisateurs et les gouvernements pour expliquer les licences et faciliter leur adoption³. Les équipes locales collaborent avec les salariés de CC pour mener des projets de recherche, émettre des suggestions en vue d'améliorer le système, faire remonter les questions, cas et problèmes des utilisateurs. Enfin, elles traduisent et produisent du matériel pédagogique et constituent un réseau pour tenter de répondre aux questions que se posent les communautés à l'échelle mondiale.

3 Hala Essalmawi, Partage de la création et de la culture : les licences Creative Commons dans le monde arabe, In : *Libres Savoirs, les biens communs de la connaissance*, C&F éditions 2012, p. 145-155.

Cependant, le processus de transposition juridique présente un inconvénient, qui provient du manque d'harmonisation entre les diverses législations du droit d'auteur. Comme les législations varient selon les pays, les licences adaptées ne recouvrent pas exactement le même champ de droits. Les licences étant déclarées compatibles entre elles, un auteur est supposé consentir à ce que les adaptations futures de ses œuvres soient licenciées selon des conditions non identifiées, ce qui peut poser un problème au regard du droit des contrats. Le processus de transposition juridique a représenté un passage constitutionnel bénéfique pour le développement d'un réseau international, et les versions adaptées facilitent la compréhension dans des systèmes et cultures juridiques différents. Mais cette activité requiert d'importants moyens humains dans un contexte de droit international complexe. Dans tous les cas, une traduction linguistique améliore l'accès, l'acceptation et la compréhension par des usagers non anglophones. Les traductions des résumés explicatifs des licences, qui condensent les principales dispositions en quelques phrases rédigées de manière non juridique, expriment clairement à tous les créateurs que les œuvres peuvent être réutilisées.

La traduction n'est pas seulement une affaire de langue. Il s'agit de rendre accessibles des concepts juridiques à un public non spécialisé ; aussi, les licences cc permettent-elles un meilleur accès à l'univers juridique, et à la connaissance et la création en général.

**LE MULTI-
LINGUISME
NUMÉRIQUE
POUR
CONSTRUIRE
DES SOCIÉTÉS
INCLUSIVES**

PARTIE 3

Près de deux milliards d'individus utilisent Internet de par le monde. Pour autant, peut-on dire que le cyberspace est accessible à tous, quelle que soit sa langue, sa culture ou ses capacités ? Comment les diasporas, les enfants, les personnes handicapées peuvent reconstruire leur univers avec l'aide du cyberspace ? Comment l'acceptation de la diversité des cultures fait bouger Internet, et comment ce réseau peut-il en retour participer au renforcement de l'inclusion sociale et mondiale ?

**VIOLA KREBS
& VICENT CLIMENT-FERRANDO**

LANGUES, CYBERESPACE ET MIGRATIONS

Comment les migrants s'emparent-ils du cyberspace ? S'expriment-ils dans leur langue maternelle ou utilisent-ils une langue véhiculaire comme le français, l'anglais ou le castillan ? En quoi leur participation contribue à la diversité linguistique dans le cyberspace et quels sont les besoins qui peuvent être partiellement ou complètement couverts pour aider les migrants dans leur nouvel environnement ? Quelles langues sont alors utilisées dans ce contexte précis ?

Article original en anglais.
Traduit par Francisca Cabrera.



VIOLA KREBS est spécialiste en matière de communication et sociolinguistique. Elle est la fondatrice et directrice exécutive d'ICVolontaires (<http://www.icvolontaires.org>), une organisation à but non lucratif œuvrant dans le domaine de la communication (technologies, culture et langues, soutien aux conférences). ICVolontaires travaille avec un réseau de 13 000 volontaires dans le monde entier. Dans le domaine de la recherche, les travaux de Viola Krebs se focalisent sur les thèmes suivants : communication, volontariat, langues et migration et éducation bilingue.



VICENT CLIMENT-FERRANDO est professeur associé à l'Universitat Pompeu Fabra de Barcelone (UPF) et conseiller gouvernemental sur les questions de langue, d'immigration et de relations internationales au sein du gouvernement de Catalogne. Il est également secrétaire général du Réseau Européen pour la Diversité Linguistique (groupe de réflexion du NPLD) et un chercheur rattaché au Groupe de Recherche Interdisciplinaire sur l'Immigration (GRITIM) au sein de l'UPF.



YOANNA RACCIMOLO a participé au présent article dans le cadre de sa collaboration à ICVolontaires. Par le passé, elle a travaillé comme chercheuse en sociologie à l'Université de Lausanne, où elle a contribué à une étude interdisciplinaire sur l'interprétation des récits, publiée dans *La pluralité interprétative enjeux théoriques et empiriques*. Très intéressée par les questions de droits humains, notamment en matière de déplacement interne et de flux migratoires, elle a par ailleurs mené des recherches en Russie et au Liban.

& VICENT CLIMENT-FERRANDO
VIOLA KRBS

ET MIGRATIONS CYBERESPACE , LANGUES

La migration n'est pas un phénomène nouveau et les statistiques montrent que la mobilité des populations dans le monde entier est en augmentation. La dernière édition de l'Atlas des Migrations¹ signale qu'il y a environ 200 millions de migrants et de personnes déplacées à travers le monde, ce qui représente environ 3 % de la population mondiale. Certains se déplacent à l'intérieur de leur propre pays, quittant le milieu rural pour les villes, d'autres quittent leur pays natal et deviennent des migrants internationaux, souvent à la recherche de meilleurs moyens de subsistance ou de conditions plus adaptées à leurs besoins quotidiens. Beaucoup de migrants ont la possibilité de quitter leur pays parce qu'ils possèdent les ressources nécessaires pour le faire, tant psychologiques qu'en matière d'éducation. Quitter sa terre natale n'est jamais une décision facile, et il y a à peine quelques mois, les vagues de migration consécutives aux insurrections en Tunisie, en Égypte et en Libye ont montré encore une fois que la migration massive peut déclencher des peurs irrationnelles et des réactions extrémistes dans les pays européens². Ainsi le facteur humain devient une impasse dans les débats politiques et économiques qui oublie la plupart du temps les origines des populations vivant dans les pays occidentaux.

« Au cours des XVII^e et XVIII^e siècles, par exil forcé, par leurre, promesses et mensonges, par enlèvement, par une nécessité impérieuse d'échapper aux conditions de vie de leur pays natal, des personnes pauvres voulant partir pour l'Amérique sont devenues des marchandises lucratives pour les

1 Wihtol de Wenden, Catherine. *Atlas des migrations dans le monde, Réfugiés ou migrants volontaires*, Alternatives Economiques, Éd. Autrement, Paris, 2009.

2 Virginie Guiraudon, directrice de recherche au CNRS et au centre d'étude européen de Sciences-po Paris
http://www.humanite.fr/01_03_2011-linvasion-de-leurope-par-bateaux-est-un-fantasme-466348

*négociants, commerçants, capitaines de navires et éventuellement pour leurs propriétaires et maîtres en Amérique*³ ». Cette description conviendrait aux peuples d’Afrique, rendus esclaves et jetés dans des embarcations pour bâtir les richesses des nations. Mais précisément les « voyous et les vagabonds » qui sont décrits correspondent à ceux dont la société Élisabéthaine voulait débarrasser ses villes anglaises. Des centaines de milliers d’entre eux, Anglais, Irlandais ou Gallois, arrivèrent en Amérique pendant cette période, devenant servants et esclaves d’autres riches immigrants d’Europe.

Même si l’œuvre de Howard Zinn, *Histoire Populaire des États-Unis*, s’attache à l’étude de l’origine du peuple américain, son exemple peut, de nos jours, être également appliqué à d’autres nations.

LANGUES EN MOUVEMENT

On estime qu’environ la moitié de la population mondiale est bilingue⁴. Plusieurs facteurs expliquent cette proportion élevée, dont les migrations de nature politique, économique et religieuse, ainsi que la politique des langues dans chaque pays.

Un grand nombre de migrants peuvent se trouver résidents allophones dans un pays dont la langue ou les langues officielles (L2) sont différentes de leur langue maternelle (L1) et de la langue ou des langues parlées dans leur pays natal. Généralement, les enfants des familles de migrants s’adaptent très rapidement à la nouvelle situation, apprennent sans difficulté la L2. Cela représente toutefois pour les parents un processus beaucoup plus lent, ceux-ci ayant souvent des grandes difficultés pour apprendre et maîtriser la langue du pays d’accueil.

Au-delà de la fonction d’échange d’information, les langues sont intimement liées à l’identité individuelle, collective et nationale. Dans ce sens, elles représentent le pouvoir, souvent la lutte pour la domination, au-delà de la question d’une simple fonction communicative. Cela ne représente rien de nouveau si l’on se réfère à l’Histoire, la situation se produisant depuis plusieurs siècles.

³ Howard Zinn, *A People’s History of the United States*, Harper Perennial Modern Classic –Persons of Mean and Vile Condition– p. 43.

⁴ Comrie Bernard et al., *The Atlas of Languages, Facts On File*, Inc., New York, USA, 1996.

L'immigration est sans doute l'un des facteurs essentiels qui conduisent aux transformations sociales, économiques et politiques. Et la question d'une gestion de la migration a été incluse dans un plus large débat, celui du multiculturalisme et des droits des minorités. En renforçant les droits fondamentaux déjà garantis à tous dans un État démocratique, le multiculturalisme permet d'amplifier la reconnaissance publique et encourage ainsi le soutien aux minorités ethniques et culturelles, afin que celles-ci puissent conserver et exprimer leurs différentes identités et pratiques culturelles⁵.

Ce qui n'empêche pas que soient soulevées des questions relatives à l'acquisition de la langue, à la présence et l'influence linguistique, aux restrictions légales établies autour des langues, ou encore au débat sur la diversité linguistique. De ce fait, dans ce monde interconnecté chaque jour davantage où la mobilité est un fait de société, l'acquisition et la maîtrise de plusieurs langues deviennent d'une importance grandissante⁶.

Cette situation complexe va de pair avec le fait que la mondialisation et l'introduction des technologies d'information et communication représentent à la fois une opportunité et une menace pour les quelque 7000 langues parlées actuellement dans le monde⁷. En effet, grâce aux TIC le monde est devenu un endroit beaucoup plus interconnecté, où l'information provenant de n'importe quel endroit du monde est transmise sur nos écrans en quelques minutes, et où les gens peuvent créer, exposer et échanger des contenus par internet.

LES RÉALITÉS DES MIGRATIONS

Le nombre de migrants est passé de 75 millions de personnes en 1965 à 200 millions en 2008, accompagnant en cela la croissance de la population mondiale laquelle a doublé pour la même période, passant de 3,2 milliards à 7 milliards⁸ de personnes. Un tiers de ces personnes migrent pour des

5 Banting, K. & Kymlicka, W (eds.). *Multiculturalism and the welfare state: recognition and redistribution in contemporary democracies*. Oxford University Press, 2006: 1.

6 Anna Lietti, *Pour une éducation bilingue*, Payot, 1994.

7 Malherbe Michel, *Les langues de l'humanité*, Robert Laffont, coll. Bouquins, 1993.
http://portal.unesco.org/es/ev.php-URL_ID=1864&URL_DO=DO_TOPIC&URL_SECTION=201.html

8 *World migration 2003: managing migration challenges and responses for people on the Move*, IOM International Organization for Migration, Geneva, 2002.

raisons familiales, un tiers sont des réfugiés et un tiers se déplacent pour des raisons professionnelles. Outre les publications de l'Organisation Internationale pour la Migration (OIM)⁹, le *CIA World Fact Book* met régulièrement à jour les données chiffrées de la migration pour chaque pays¹⁰.

On assiste depuis vingt ans à une augmentation de la migration dans le sens Sud-Sud. Ainsi, l'Asie représente le plus grand nombre de migrants, avec 40 à 50 millions de migrants Chinois et 20 millions de migrants Indiens.

En outre, les mouvements migratoires ont atteint une complexité et une étendue sans précédent dans l'histoire. Aux pays d'accueil traditionnels comme les États-Unis et le Canada, il faut désormais ajouter des pays qui – il n'y a pas très longtemps encore – étaient des pays d'émigration comme l'Espagne, l'Italie, le Portugal, entre autres. Il y a ainsi aujourd'hui des groupes d'immigrants dans quasiment toutes les démocraties occidentales.

Dans l'Union Européenne, les pays qui ont reçu plus de la moitié des immigrants en 2008 sont l'Espagne, l'Allemagne et le Royaume-Uni¹¹.

Un exemple très représentatif est celui du phénomène de migration en Catalogne. En 1998, seulement 1.97 % des 6 147 610 habitants de la Catalogne étaient des immigrants étrangers. En 2009, la population avait augmenté de plus d'un million de personnes dont au moins 17 % étaient immigrants. De ce fait, la Catalogne est la région qui reçoit le plus grand nombre d'immigrants dans toute l'Espagne. De surcroît, l'arrivée de personnes d'origine étrangère représente 77 % de la croissance moyenne de la population dans cette région pour la même période.

À Genève, en Suisse, 38.3 % des résidents sont étrangers. 184 nationalités y sont représentées, parlant environ 150 langues¹². Pour 25 % de la population habitant à Genève, le français n'est pas la langue principale et une partie d'entre eux n'est pas capable de le comprendre et/ou de le

9 <http://www.iom.int>

10 <http://www.cia.gov/library/publications/the-world-factbook/fields/2112.html>

11 Source : Eurostat, European Statistics available online at http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Migration_statistics figures from 7 November 2010.

12 Portrait statistique des étrangers vivant à Genève. Résultats du recensement fédéral de la population et autres sources, *Office cantonal de la Statistique, Études et documents* n° 37, Genève, septembre 2005.

parler¹³. Rien qu'à Londres on a recensé 300 langues différentes. Le dernier recensement au Canada présente le chiffre de 6 293 110 allophones, ce qui veut dire que 20,1 % de la population totale du pays parle une autre langue en plus de l'anglais et du français. Les États-Unis sont confrontés au même phénomène : le recensement de l'année 2000 a fait état de plus de 300 langues parlées dans le pays¹⁴.

Même si des villes comme Barcelone, Londres, New York et Genève sont particulièrement internationales, cette montée du multiculturalisme est une tendance générale. Si l'on suppose que chaque citoyen a le droit d'accéder aux soins de santé et à l'éducation, il nous faudrait examiner quelles langues sont utilisées pour communiquer dans ces contextes dans un cas spécifique. Il y a un risque d'exclusion sociale et professionnelle de la population allophone.

POLITIQUES ET THÉORIES DE L'IMMIGRATION

Longtemps, des pays comme l'Australie, le Canada et les États-Unis – principaux pays d'accueil – ont adopté une stratégie d'assimilation des immigrants qui étaient censés s'intégrer complètement à la société dominante. On espérait ainsi, au bout d'un certain temps, assister à l'effacement des différences d'avec les citoyens natifs et leur façon de vivre.

La fin des années soixante et le début des années soixante-dix virent deux changements majeurs : premièrement, l'adoption de critères d'admission neutres au regard des aires d'origines, géographiques et culturelles, dans le but d'éviter l'exclusion des immigrants issus de sociétés non européennes (et souvent non chrétiennes) ; deuxièmement, l'adoption d'un concept d'intégration davantage « multiculturel », censé permettre l'expression des identités ethniques, secondé par l'obligation des institutions publiques de s'accommoder de ces identités¹⁵. Il est cependant nécessaire de souligner que cette tendance n'a pas pris forme dans toutes les sociétés occidentales et que le degré de reconnaissance publique varie selon les pays.

13 *La politique cantonale de préformation des non-francophones à risque d'exclusion : Evaluation des mesures de soutien*, Commission externe d'évaluation des politiques publiques, Genève, Septembre 2005, page 15.

14 <http://www.census.gov/prod/2003pubs/c2kbr-29.pdf>

15 Banting, K. & Kymlicka, W (eds.). *Multiculturalism and the welfare state : recognition and redistribution in contemporary democracies*. Oxford University Press, 2006 : 54.

En sus de traditions, religions et cultures différentes, les immigrants ont apporté avec eux un tout nouvel éventail de langues. Par exemple, le tamazigh – parlé dans des régions éparses du Maroc, d’Algérie, de la Tunisie, de l’Égypte et du Mali – est la troisième langue la plus parlée en Catalogne, après le catalan et l’espagnol.

Ce changement sans précédent dans le kaléidoscope linguistique des sociétés occidentales ne semble pas pour autant avoir conduit à d’intenses débats académiques concernant la façon d’appliquer les politiques multiculturelles aux langues des migrants, et on ne trouve que peu de références sur la question dans la littérature scientifique.

Dans le cadre d’une réflexion multiculturelle sur les langues des migrants, la question la plus souvent soulevée est celle de leur éducation dans leur langue maternelle¹⁶ et donc de l’instauration de la langue appartenant à certains groupes spécifiques de migrants en tant que langue véhiculaire pour des activités extrascolaires. Des études scientifiques portant sur le bilinguisme ou le plurilinguisme montrent que l’éducation bilingue est une bonne manière de transmettre plusieurs langues et que son apport au développement cognitif de l’enfant est positif¹⁷.

L’éducation bilingue ou l’enseignement de la langue maternelle sont deux formes d’instruction adoptées dans les années soixante-dix par un certain nombre de pays (Allemagne, Pays-Bas, etc.) lorsqu’il était généralement admis que les migrants repartiraient chez eux. À l’origine, ce n’était donc pas la conception d’une politique multiculturelle. En outre, l’éducation bilingue aux Pays-Bas et en Allemagne a récemment reculé, étant parfois considérée comme un obstacle à l’assimilation. De manière générale, la reconnaissance de la langue des migrants commence à être questionnée comme allant à l’encontre de l’efficacité de leur intégration¹⁸. En effet, l’argument mis en avant est que la reconnaissance et le soutien aux langues

16 Voir dans cet ouvrage : Marcel Diki-Kidiri, *Le cyberspace dans l’éducation en langue maternelle*.

17 Cummins Jim. *Bilingualism, multiculturalism, and second language acquisition: the McGill conference in honor of Wallace E. Lambert*. Ed. L. Erlbaum, London, 1976, 1981.

CDIP, Conférence suisse des directeurs cantonaux de l’instruction publique. *Rapport sur la question des langues, Quelles langues apprendre en Suisse pendant la scolarité obligatoire ?* Berne, 1998.

Crawford James. *Bilingual Education: History, Politics, Theory and Practice*. Bilingual Educational Service, Inc., Los Angeles, 1999.

18 Kymlicka, W & Patten, A. (eds). *Language Rights and Political Theory*. Oxford University Press, 2003 : 8-9.

des migrants pourraient déboucher sur des ghettos et des enclaves linguistiques, ce qui serait nuisible à la fois pour la société d'accueil – qui voit l'enseignement de sa langue comme source de patriotisme et de loyauté – et pour les groupes d'immigrants – dont le manque d'aisance dans la langue du pays d'accueil pourrait mener à une marginalisation et à une précarité économique. L'un des mots-clés de ce débat est « société parallèle », soit la peur de promouvoir les particularités des immigrés – cultures, religions et langues – et de créer par là des communautés ethniques closes et séparatistes, conduisant à ce que la littérature spécialisée dénomme la « balkanisation ».

L'accent semble porter dorénavant sur les langues des sociétés d'accueil. Le nombre de propositions qui conditionnent la naturalisation aux tests de langues est en augmentation. Ce genre de propositions émerge dans tous les pays de l'Occident où les difficultés d'intégration des immigrants sont souvent attribuées à l'incapacité ou à la mauvaise volonté des immigrés à apprendre la langue officielle du pays. Dans certains pays d'Europe, il est même question d'instaurer une loi obligeant les immigrants à suivre des cours de langue, condition préalable pour accéder aux aides sociales. Une étude récente¹⁹, analysant les programmes appliqués en Allemagne, France, Canada et les Pays-Bas, confirme cette tendance. L'obligation faite aux migrants d'apprendre la langue du pays de résidence peut être interprétée comme un « retour à l'assimilation ».

Un autre exemple de la façon dont les politiques exercent un impact sur les migrations est donné par la politique de la « Porte Ouverte » établie par le gouvernement chinois dans les dernières décennies, dont on retrouve l'impact dans les statistiques des migrations. Dans un article récent, Phœbe H. Li décrit ces changements²⁰. En 1997, le nombre total de citoyens chinois partant à l'étranger était de 5,32 millions. Dix ans plus tard, en 2007, il y eut environ 80 millions de passages frontaliers dans toute la Chine continentale. Ces ressortissants chinois sont partis vers un grand nombre de destinations dans le monde et comprenaient un large éventail de migrants permanents ou temporaires, tels des étudiants, travailleurs sous contrat ou touristes, parmi lesquels une grande

19 Biles, J et al. Policies and Models of Incorporation. A Transatlantic Perspective : Canada, Germany, France and the Netherlands. *Documentos Cidob. Migraciones*, num. 12, June 2007.

20 Phœbe H. Li, University of Auckland, New Zealand, *New Chinese Immigrants to New Zealand, a PRC Dimension*, 2010, http://international.metropolis.net/pdf/fow_newzealand_immi.pdf

partie étaient de potentiels migrants permanents. Phœbe H. Li décrit les critères d'immigration appliqués en Nouvelle Zélande, dont notamment leurs capacités à s'intégrer professionnellement et leurs compétences en langue anglaise.

Plus récemment, des ressortissants chinois sont partis massivement vers l'Afrique. Beaucoup d'entre eux travaillent dans le secteur de la construction, tandis que d'autres se sont lancés en affaire, créant une compétition inopportune au commerce local, dans le secteur du textile, de la construction de routes, de la restauration, etc. Dans certains cas, ils créent également des activités commerciales dans le secteur informel. D'un point de vue linguistique, beaucoup de ces nouveaux immigrants ne parlent pas les langues locales. De ce fait, ils tendent à vivre entre eux et de façon générale ne s'intègrent pas aussi facilement que les migrants venus d'autres régions du monde²¹.

Il existe une catégorie de migrants qui restent en retrait de la société locale et sont dépendants d'une ou deux personnes uniquement pour établir le lien avec la langue du pays d'accueil. Les immigrants du Sri Lanka au Liban sont l'exemple parfait de ce scénario. La plupart sont employés en tant que domestiques et la grande majorité d'entre eux ne sont pas autorisés à quitter la maison sans le propriétaire, qui d'habitude confisque leur passeport en les obligeant à vivre d'une façon isolée, proche de l'esclavage. Pour éviter cet isolement, des ONG comme *Caritas Migrant*, apportent à ces populations un soutien légal et administratif²².

Les exemples cités montrent la complexité de la situation concernant les politiques d'immigration et ses enjeux. La solution aux problèmes linguistiques dépend en grande partie de la façon dont les gouvernements se comportent face à l'immigration. Mais ils dépendent aussi de deux autres critères importants.

Le premier concerne le type de migration, qui peut être saisonnier, transitoire ou circulaire. Dans tous ces cas, les migrants séjournent quelques années dans un autre pays, puis retournent dans leur pays d'origine ou se déplacent ailleurs. L'intégration linguistique n'est habituellement pas considérée comme une priorité car leur objectif principal est de gagner de l'argent ou de construire une carrière plutôt que de s'intégrer dans le

21 *Speak to me, speak here. Situation linguistique à Barcelone.* Éd. ICVolontaires.org. 2007.

22 Caritas Lebanon Migrant's Center : <http://www.caritas.org.lb>

nouveau pays et société. En conséquence, beaucoup de ces migrants ne sont pas intéressés à apprendre la langue du pays d'accueil temporaire.

Le deuxième critère est plutôt lié à certains aspects psychologiques des migrants, pouvant occasionner des obstacles pour l'apprentissage de la langue nationale. Beaucoup décident de ne pas faire des efforts pour apprendre la langue. Ce refus symbolise leur intention personnelle de ne pas s'établir pour toujours dans un pays étranger. Ils maintiennent un désir profond de rentrer chez eux et ne voient pas l'intérêt d'apprendre une langue pour un séjour considéré comme temporaire. En réalité, beaucoup d'entre eux y séjournent souvent longtemps sans acquérir une compréhension significative de la langue du pays d'accueil car ils ne veulent pas renoncer à l'espoir de rentrer chez eux.

INTERNET COMME UNE OPPORTUNITÉ POUR LES MIGRANTS : QUELQUES EXEMPLES D'OUTILS ET RÉSEAUX DÉJÀ EXISTANTS

L'Internet et le Web offrent des ressources et des outils importants aux migrants ainsi qu'aux travailleurs sociaux, traducteurs, interprètes et décideurs, dont le travail a un lien avec les questions soulevées par l'immigration.

En premier lieu, les migrants peuvent apparaître sur le web en tant que sujet d'articles. Beaucoup de journalistes écrivent sur les migrants et sur leur impact au sein des pays qui doivent parfois gérer des flux d'immigration massive.

En deuxième lieu, des ressources en ligne sont créées dans l'intention d'aider les migrants dans leur nouvel environnement. Certaines sont disponibles dans la langue du pays d'accueil, d'autres sont traduites dans les langues parlées par les populations de migrants provenant de tel ou tel pays.

En troisième lieu, les migrants utilisent le Web pour garder contact avec leur famille distante, avec leur pays et leur culture d'origine. Cela peut se passer aussi bien à travers des vidéos ou des communications téléphoniques, qu'au travers de la lecture des contenus de sites publiés dans la langue locale de leur pays ou de leur région.

Sites de recherche sur les migrations

Pour mieux comprendre la question des mouvements migratoires, plusieurs sites de recherche mettent à disposition des articles et des études sur le sujet. Le MPI²³ (*Migration Policy Institute*) est un *think tank* indépendant à but non lucratif, dédié à l'étude des mouvements des populations à travers le monde. Sa base de données en ligne met à disposition un grand nombre d'articles et présente des tableaux approfondis des plus récentes données concernant les tendances migratoires et ses modalités aux États-Unis et dans le monde. Les outils de recherche incluent une base de données rapportant le nombre des naissances d'enfants étrangers aux États-Unis, État par État (*US State Data on the Foreign Born*), des cartes des naissances (*Maps of the Foreign Born*), la carte de la migration mondiale (*The World Migration Map*), des graphiques et tableaux comparatifs, le guide mondial des versements et envois d'argent à l'extérieur (*Global Remittances Guide*) et des données concernant l'asile.

Un autre groupe de recherche intéressant est le « Programme d'études sur l'usage des TIC (*Technologies de l'Information et de la communication*) dans les migrations » (TIC-Migrations)²⁴. Il se définit comme un programme « explorant l'impact des nouvelles technologies sur le monde des migrants » (trajectoires, connexions personnelles, relations avec le pays d'origine et le pays d'accueil, etc.). Ses objectifs sont « d'ouvrir un nouveau terrain de recherche, articuler deux courants d'études jusqu'ici relativement dissociés (*la théorie des diasporas et l'exploration du web*) et développer des outils génériques à redéployer dans d'autres disciplines des Sciences Humaines et Sociales ». Il présente aussi le concept de « migrants connectés » qui essaie de cerner la figure du migrant dans un monde fait d'une continuité discontinue.

Un autre site présentant des études liées aux questions de la migration est le Projet International de *Metropolis*²⁵, qui est un forum qui compte créer des liens entre études, politiques et pratiques touchant les migrations et ses difficultés. Ce projet a pour but d'accroître la capacité de recherche académique, d'encourager les travaux de recherche pertinents et de rendre plus facile leur utilisation par les organisations gouvernementales ou non-gouvernementales. Les articles sont disponibles en ligne et les conférences

23 <http://www.migrationinformation.org>

24 <http://ticmigrations.fr>

25 <http://international.metropolis.net>

de Metropolis réunissent chaque année plusieurs centaines de chercheurs impliqués dans une recherche universitaire sur les migrations.

Le dernier projet que nous pouvons citer ici s'appelle *Bridge-IT Network*²⁶ qui soulève la question du potentiel des TIC pour favoriser l'intégration des migrants et la diversité culturelle en Europe.

Ressources d'information aidant les migrants dans le pays d'accueil

Les sites dédiés aux migrants apportent souvent des informations pratiques et des liens utiles pour la vie de tous les jours dans le pays de résidence. Un exemple de ce genre de sites est la page suisse Migraweb²⁷, qui présente des informations pratiques rassemblées par des organisations indépendantes de la société civile, mais aussi par des autorités publiques compétentes et par des spécialistes dans chaque domaine. Ces informations sont ensuite traduites vers les langues parlées des migrants par un groupe de volontaires qui sont eux-mêmes des migrants établis et appartenant à diverses communautés. Ces volontaires parlent couramment leur langue d'origine, ainsi que le français ou l'allemand. Ils fonctionnent ainsi comme un pont entre les différentes cultures.

En Espagne, MIGRAweb.es²⁸ permet l'accès à une équipe de professionnels donnant des conseils juridiques concernant particulièrement les problèmes relatifs à l'immigration, l'asile et la nationalité.

MigraLingua.org²⁹ cherche à apporter des renseignements pratiques sur les services d'interprètes communautaires coordonnés par ICVolontaires, qui ont pour but d'accompagner les migrants ne parlant pas le français dans leur vie quotidienne à Genève. L'objectif est d'augmenter la compréhension réciproque entre migrants et professeurs, d'encourager un dialogue respectueux, rendre plus facile l'accès aux institutions, aider les migrants à sortir de leur isolement et de ce fait leur donner l'envie d'apprendre le français pour leur procurer une plus grande indépendance. Grâce à l'apport d'un réseau de volontaires, ce service est disponible à toute personne quelles que soient ses origines et sa situation sociale.

²⁶ <http://bridge-it.ning.com>

²⁷ <http://www.migraweb.ch>

²⁸ <http://www.migraweb.es>

²⁹ <http://www.migralingua.org>

Il existe également un grand nombre de logiciels développés en vue de combler la distance linguistique entre patients et médecins dans le système de santé. L'*Universal Doctor Speaker Project*, un logiciel développé par un groupe de médecins, mérite d'être mentionné dans ce contexte. Ce projet est mis en œuvre dans plusieurs hôpitaux européens et a pour but de faciliter la communication entre les professionnels de la Santé et les patients provenant de différents milieux. Il est disponible en dix langues, dont notamment l'ourdou et le mandarin. Des applications relatives à ce projet existent pour iPhone et iPad³⁰.

Les téléphones mobiles peuvent également fournir un appui aux migrants. Le projet *Voix Mobiles (VozMob)* est un des meilleurs exemples, le téléphone étant utilisé pour transmettre des témoignages, procurer du soutien aux membres d'une même communauté, et leur permettre de garder le contact avec la famille ou les compatriotes. *VozMob* se décrit lui-même comme étant « *une plateforme permettant aux immigrants et/ou aux travailleurs à très faible revenu de Los Angeles et de sa région de créer des récits de leurs vies et de leurs communautés directement à partir des téléphones mobiles* »³¹.

Cours de langue en ligne

Désormais, il est possible d'apprendre une langue par internet. Des cours payants en ligne, des forums de discussion et des programmes d'éducation sont accessibles sans peine pour ceux qui souhaitent apprendre ou améliorer certaines des différentes langues parlées dans les pays d'accueil. Un nombre grandissant de sites web offre un large choix de leçons payantes dans une quarantaine de langues environ : anglais, allemand, espagnol, italien, russe, arabe, chinois (mandarin), farsi, hindi, etc.³² Cependant, les langues les plus représentées sont l'anglais, le français, l'espagnol et l'italien. Ce genre de cours peut être une excellente alternative lorsqu'il est difficile au migrant de suivre de manière régulière des cours de langue compatibles avec les horaires et les contraintes de son travail.

30 <http://itunes.apple.com/us/app/universal-doctor-speaker-for/id364812043?mt=8>

31 <http://vozmob.net/en/about>

32 <http://www.livemocha.com>

QUAND LA TECHNOLOGIE DEVIENT UN DÉFI PLUTÔT QU'UNE SOLUTION

De manière générale, les TIC connectent les personnes se trouvant dans différentes régions géographiques. Cependant, ces technologies peuvent être un véritable problème pour les plus démunis, souvent exclus de l'accès aux technologies informatiques. En effet, la technologie a un prix : équipements, souscription à internet et les cybercafés ne sont pas encore abordables pour tout le monde. Cela s'applique aux immigrants mais aussi aux proches restés sur place. En termes de budget, une partie des migrants vivant dans les pays industrialisés ne se trouve pas à même de pouvoir utiliser des ordinateurs et internet.

Plusieurs études démontrent l'impact positif des TIC sur les échanges des migrants avec leur famille et leur communauté. Ces études montrent notamment que la « culture du lien »³³ est très souvent encouragée, et souvent les services internet tels que Skype ou MSN Messenger, ainsi que la possibilité d'utiliser des webcams, sont tellement loués qu'on n'en retient que les aspects favorables. Même si les migrants peuvent garder le contact avec leurs proches et les voir très souvent grâce aux webcams, ces pratiques peuvent aussi provoquer des effets contraires. Au lieu d'avoir l'impression d'être proche de la famille et des amis, le migrant peut par exemple ressentir la nostalgie de ne pas être physiquement présent pour un anniversaire ou un événement spécial. De plus, dans tous les cas, la communication doit être interrompue et elle ou lui est « condamné » à rester séparé par la distance. Si les TIC, certes, peuvent procurer une amélioration, ils ne remplacent pas pour autant les rapports personnels directs. Idéaliser les TIC, c'est sous-estimer les difficultés réelles rencontrées par les migrants séparés géographiquement de personnes aimées.

Un autre aspect négatif de tels changements réside dans une intensification des obligations des membres de la famille ayant migré de « maintenir le contact », ce qui peut provoquer une intense pression émotionnelle. En effet, la plupart du temps, les proches restés dans le pays d'origine idéalisent les conditions de vie des membres de la famille partis à l'étranger et sont convaincus qu'ils gagnent beaucoup d'argent, ce qui peut ou non s'avérer vrai. La pression de devoir envoyer de l'argent chez soi peut être importante et elle a très probablement augmenté ces dernières années

33 <http://ticmigrations.fr/fr/whoweare/manifeste>

grâce à la diversité des moyens de communication disponibles de nos jours.

Un autre problème concerne la relation entre le migrant et l'administration. Aujourd'hui plus nombreux sont les services en ligne, administratifs ou privés, qui remplacent ce qui était autrefois assuré par le service public ou les agences de service aux clients, dans lesquels une personne physique se trouvait en situation de répondre à l'usager. Par exemple, en France, le site web officiel de l'Agence nationale pour l'emploi³⁴, requiert une inscription en ligne pour pouvoir bénéficier des allocations chômage. Ces sites sont souvent disponibles uniquement dans la langue locale (dans le cas cité, le français), ce qui n'est pas vraiment d'une grande aide aux allophones. Dans ce contexte, les services administratifs réduisent en réalité leurs coûts et leurs responsabilités en interposant une machine entre les demandeurs de ces services et les employés qui traitent leurs demandes.

L'administration électronique a également un grand impact sur l'identification. Les systèmes biométriques d'identification deviennent de plus en plus fréquents et acceptés. Ils renforcent la société de contrôle, et règlent plus étroitement l'afflux des immigrants³⁵. Les migrants sont soumis à des contrôles intensifs motivés par la peur des migrations illégales.

En outre, les migrants peuvent être illettrés ou peuvent n'avoir aucune expérience d'internet. Ils sont dès lors totalement dépendants d'amis ou d'autres membres de leur famille s'ils sont présents, parfois même de leurs enfants qui vont à l'école et savent utiliser ces nouvelles technologies ainsi que parler la langue du pays d'accueil. Les deux premières situations créent des relations asymétriques qui peuvent affecter négativement les personnes en relation. Les amis sollicités sont obligés de se mêler d'affaires privées, alors qu'ils préféreraient sans doute ne pas avoir à connaître. Mais les amis n'ont pas toujours le courage de refuser de l'aide dans des cas d'absolue nécessité. Du côté du migrant, une dépendance vis-à-vis des personnes qui n'appartiennent pas au cercle familial se développe et elle, ou lui, doivent partager des informations parfois très intimes.

Les enfants, quant à eux, se trouvent souvent à faire face à des situations où ils doivent effectuer des démarches administratives, en personne ou en ligne. Ce point représente un des défis soulignés par une étude

34 <http://www.pole-emploi.fr>

35 <http://ticmigrations.fr/fr/whoweare/manifeste>

récente menée à Genève sur la situation du soutien linguistique fourni aux migrants³⁶. Des témoignages montrent que ces enfants, étant les seuls à avoir une certaine compréhension de la langue locale, souffrent grandement, parfois en silence, supportant presque exclusivement le poids des problèmes de la migration et de l'adaptation culturelle. Des nombreuses études montrent que ces enfants d'immigrants sont très tôt sujets à des pressions psychologiques provoquant un autisme volontaire causé par le sentiment de déloyauté envers leurs parents. Ils sont déchirés entre deux espaces ; le premier, l'espace de leur maison, leur langue maternelle et leur famille, et le second, l'univers hostile de la société inconnue et la langue étrangère que leurs parents ne maîtrisent pas. La seule solution qu'ils trouvent est de se murer dans le silence, comme pour établir une barrière entre ces deux mondes antagonistes³⁷.

Les deux attitudes de silence précitées s'alimentent mutuellement et rendent inévitablement encore plus difficile la situation de l'enfant, entraîné dès lors dans un cercle vicieux.

RECOMMANDATIONS / OPPORTUNITÉS

Si un tel article ne peut jamais être exhaustif, il montre que migrations et problèmes linguistiques sont dans une interaction complexe qui touche plusieurs millions de personnes dans le monde. La technologie peut être un atout pour les migrants, notamment quand ils vivent dans un milieu dont la ou les langue(s) dominante(s) sont différente(s) de celle(s) qu'ils comprennent ou parlent. Les traducteurs en ligne ainsi que les autres ressources numériques peuvent être d'une grande aide. Pourtant, le fait qu'un nombre grandissant d'administrations demande un enregistrement en ligne aux citoyens demandeurs de leurs services représente un défi surtout pour les allophones et les illettrés.

Le rôle des enfants dans ce contexte est particulièrement délicat et doit être pris en considération. Les enfants ne doivent pas remplacer les travailleurs sociaux ni être placés dans des situations où ils doivent assumer les rôles des adultes.

36 Viola Krebs et al., *Voix au Chapitre, Rapport sur l'accompagnement linguistique des migrants non-francophones à Genève*, Éd. ICVolontaires.org, 2008.

37 Dahoun Zerdalia, *Les couleurs du silence. Le mutisme des enfants migrants*, Calmann-Lévy, 1995.

Des efforts comme ceux ayant abouti à la *World Charter of Migrants*³⁸ aident à circonscrire le problème et proposent des solutions en vue de développer des politiques d'immigration décentes et pratiques. Dans la même lignée, une liste concise de conventions et de déclarations rapportées aux politiques d'immigration est disponible sur le site de l'Organisation Internationale pour les Migrations.

Le rôle des outils et des sources d'information en ligne est de partager l'information, de fournir une assistance et de promouvoir l'usage de langues autres que celles construisant les débats internationaux. Mais ce sont des outils et des guides dont le développement et la traduction exigent en premier lieu un investissement humain.

BIBLIOGRAPHIE

[ANAYA JAMES 1996] Anaya James. *Indigenous Peoples in International Law*. New York: Oxford University Press, 1996.

[BAUBÖCK] Bauböck R. "Cultural citizenship, minority rights and self-government". In : T. Aleinikoff and D. Klusmeyer (eds) *Citizenship Today: Global Perspectives and Practices*. Washington : Carnegie Endowment for International Peace.

[BRUBAKER ROGERS 2003] Brubaker Rogers. "The Return of Assimilation? Changing Perspectives on Immigration and its Sequels in France, Germany and the United states", In : Joppke, Christian/Morawska, Ewa (eds.), *Toward Assimilation and Citizenship. Immigrants in Liberal Nation-States*, New York, 2003.

[CLIMENT-FERRANDO 2011] Climent-Ferrando, Vicent « La recherche sur l'immigration en Catalogne : bilan 2000-2010 » [Research on Migration in Catalonia : An Overview 2000-2010] In : *Migrations et Société. Centre d'études sur les migrations internationales*, vol XXIII, 134-135, Paris, Mai-Juin 2011, 251-268.

[EUROSTAT] Eurostat, European Statistics available online at http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Migration_statistics, figures

[JOPPKE 2004] Joppke Christian. "The retreat of multiculturalism in the liberal state: theory and policy" *The British Journal of Sociology*, Volume 55/2 ; 237-257, 2004.

[JUNYENT 2005] Junyent Carme (ed.). *Les llengües a Catalunya. Quantes llengües s'hi parlen?* Barcelona, Editorial Octaedro, 2005.

[KREBS 2007] Krebs Viola. Bilinguisme, inter culturalité et communication politique, dans *L'anglais et les cultures : carrefour ou frontière ? Droit et Cultures*, éd. Harmattan, Paris, 54/2007, <http://droitcultures.revues.org/79>

[KYMICKA, NORMAN 2000] Kymlicka, W. & Norman, W. *Citizenship in diverse societies*. Oxford, Oxford University Press, 2000.

³⁸ <http://www.cmmigrants.org>

[KYMILICKA 2001] Kymlicka, W. *Politics in the vernacular: nationalism, multiculturalism, and citizenship*. Oxford. Oxford University Press, 2001.

[KYMILICKA 1995] Kymlicka, W. *Multicultural citizenship: a liberal theory of minority rights*. Oxford, Clarendon, 1995.

La politique cantonale de préformation des non-francophones à risque d'exclusion : Évaluation des mesures de soutien, Commission externe d'évaluation des politiques publiques, Genève, Septembre 2005, p. 15.

[MCROBERTS 2001] McRoberts, Kenneth. "Canada and the Multinational State", *Canadian Journal of Political Science*, 683-714, 2001.

[ZAPATA-BARRERO 2007] Zapata-Barrero Ricard. "Immigration, Self-Government and Management of Identity, The Catalan Case". In: Korniman, M. ; Lauglanf, J. *The Long March to the West: 21st Century Migration in Europe and the Greater Mediterranean Area*. Vallentine-Mitchell, 2007.

[ZAPATA-BARRERO 2007] Zapata-Barrero Ricard. "Setting a research agenda on the interaction between cultural demands of immigrants and minority nations" *Journal of Immigration and Refugee Studies* vol.5, n°4; 1-25, 2007.

Sites web

BABELFISH <http://babelfish.yahoo.com>

CATALAN TRANSLATOR http://traductor.gencat.net/index_en.jsp

EUROSTAT. STATISTIQUES DE LA COMMISSION EUROPÉENNE http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/migration_statistics

GOOGLE AUTOMATIC TRANSLATOR <http://translate.google.com>

INSTITUT STATISTIQUE CATALAN <http://www.idescat.cat/dequavi/dequavi.?tc=444&v0=1&v1=8>

INSTITUTO NACIONAL DE ESTADÍSTICA. GOBIERNO DE ESPAÑA http://www.ine.es/inebmenu/mnu_cifraspob.htm

MAIRIE DE BARCELONE, CATALOGNE <http://www.bcn.cat>

METROPOLIS <http://international.metropolis.net>

MIGRALINGUA <http://www.migralingua.org>

MIGRATION INFORMATION <http://www.migrationinformation.org>

MIGRATION ONLINE <http://www.migrationonline.cz/themes/eu/>

MIGRAWEB <http://www.migraweb.ch>

MIGRAWEB.ES <http://www.migraweb.es>

TRANSLATION CAFÉ <http://translatorscafe.com>

ACCESSIBILITÉ DU CYBERESPACE : LANGUES DES SIGNES

Les langues des signes sont les langues naturelles des communautés de personnes sourdes. Leurs lexiques et grammaires sont très différents de ceux des langues vocales. Dans de nombreux pays, la langue vocale, même sous sa forme écrite, est généralement une seconde langue souvent mal maîtrisée par les locuteurs sourds. Le traitement automatique des langues des signes est un domaine très récent. Dans cet article, nous faisons un état des lieux des ressources existantes, tant en terme de corpus que de systèmes et outils basés sur les nouvelles technologies et les usages associés.

Article original en français.



ANNELIES BRAFFORT est directrice de recherche au LIMSI-CNRS et coordonne les recherches en Traitement Automatique des Langues des Signes (TALS). Ses recherches concernent l'étude de corpus, la modélisation du fonctionnement de la Langue des Signes Française (LSF) et la traduction du texte vers la LSF. Les applications actuelles portent sur la génération automatique par l'animation de signeurs virtuels, personnages 3D s'exprimant en LSF.



PATRICE DALLE est professeur de l'Université Toulouse 3. Il dirige l'équipe de traitement et compréhension d'images de l'IRIT. Ses travaux portent sur la modélisation, le traitement et la compréhension des Langues des Signes. Les applications concernent la communication en langue des signes et le développement d'outils pédagogiques.

ANNELIES BRAFFORT
& PATRICE DALLE

ACCÉSSIBILITÉ DU CYBERESPACE : LANGUES DES SIGNES

Les langues des signes (LS)⁴ sont les langues naturelles utilisées par les sourds et certains entendants côtoyant des locuteurs sourds. Ces langues, dites visuo-gestuelles (car émises par le corps et reçues via la vision), s'expriment dans l'espace en face du locuteur, au moyen d'unités gestuelles composées de gestes des mains et des bras, de mouvements du buste, des épaules et de la tête, d'expressions du visage, de directions du regard, etc., réalisés simultanément.

Les LS représentent le seul type de langue à disposition des sourds profonds pour communiquer avec l'entourage, les langues vocales (LV) ne pouvant pas être perçues. Pour les sourds, les langues signées ne sont que très minoritairement des langues maternelles : la surdité est héréditaire dans seulement 4 % des cas. Ainsi, pour 96 % des locuteurs sourds, la transmission de la langue est a priori non parentale². Les LS assurent toutes les fonctions remplies par les autres langues naturelles non gestuelles et sont, pour les sourds, le seul mode linguistique véritablement approprié, le seul qui leur permette un développement cognitif et psychologique d'une manière équivalente à ce qu'il en est d'une langue vocale pour un entendant [DALLE 2005].

De même que pour les LV, il n'existe pas une LS universelle, mais autant que de communautés différentes de sourds, chaque LS ayant son histoire, ses unités signifiantes et son lexique³. Cependant, deux sourds maîtrisant chacun une LS différente parviennent à se comprendre et à communiquer dans un délai très bref, sans commune mesure avec ce qu'il en est pour

1 Tout au long de cet article nous utiliserons les abréviations : LS pour Langues des Signes et LV pour Langues Vocales.

2 <http://corpusdelaparole.in2p3.fr/spip.php?article117>

3 Lorsque nous parlerons de LSF et de français, ces propos peuvent être généralisés à la LS d'un pays donné et de la LV de ce pays.

deux locuteurs de LV différentes. Ceci tient à la grande proximité des structures linguistiques existant entre les diverses LS dont certaines, très iconiques, sont caractérisées par l'absence de signes dits « standard »⁴ (qui sont eux différents pour chaque langue) [CUXAC 2000]. L'origine de ces structures partagées tient probablement à la nature du canal et à sa proximité avec les représentations mentales des locuteurs sourds. On parle souvent d'une « pensée visuelle ». Elle remet en question ce que nous considérons habituellement comme appartenant au domaine de la linguistique.

Seules quelques-unes parmi la centaine de langues des signes dans le monde ont obtenu une reconnaissance légale, les autres ne bénéficiant d'aucun statut officiel. En France, la Loi n°2005-102⁵ *Pour l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées* reconnaît officiellement la Langue des Signes Française (LSF) depuis 2005.

Les LS sont éminemment orales parce que sans écriture. Ainsi les LS se différencient doublement des langues institutionnelles dans la mesure où, gestuelles, elles ne sont pas vocales et où elles sont uniquement orales car sans systèmes d'écriture. Les LS sont donc des langues orales sans autre recours de transmission et d'enseignement que le face-à-face immédiat ou différé à l'aide de vidéos.

Une des raisons essentielles expliquant la difficulté à créer ou à emprunter une écriture pour les LS réside dans la modalité de ces langues : elles s'inscrivent dans l'espace et dans le temps par des gestes, des postures, des mimiques et des jeux de regards, tous signifiants et potentiellement simultanés. Le mode d'expression de ces langues est donc multilinéaire et pluridimensionnel. Or l'ensemble des écritures que l'homme a forgées suivent la monolinéarité des langues vocales. Actuellement, aucune technologie graphique ne permet de remplir, pour les LS, les fonctions quotidiennes et primaires de l'écrit (trace, notation à la volée, lecture linéaire), même si la vidéo est parfois utilisée pour remplir certains de ces rôles [BRUGEILLE 2006]. La seule écriture dont disposent les sourds est généralement celle de la LV du pays où ils vivent. Mais la majorité des sourds profonds ne maîtrisent pas assez bien l'écrit pour leur permettre

4 Unité lexicale institutionnalisée pour une LS donnée, éventuellement répertorié dans un dictionnaire.

5 <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000809647>

d'accéder à des études et à des formations d'un niveau élevé, d'avoir accès aux moyens de communication basés sur l'écrit ou encore d'assumer leur rôle de citoyen, freinant ainsi leur vie professionnelle mais aussi personnelle.

Le fait que les LS soient des langues sans écriture et que l'accès à l'écrit des LV soit difficile pour la grande majorité des sourds profonds a pour conséquence une accessibilité très incomplète aux informations du cyberspace. Si quelques rares sites proposent une traduction des écrits en LS à l'aide de vidéo, cela reste très limité et la mise à jour n'est pas évidente.

De ce fait, il est intéressant de mener des recherches en traitement automatique des langues des signes (TALS), sur les trois aspects : reconnaissance, génération et traduction automatique. Ces travaux de recherches s'appuient sur l'analyse de corpus, principalement vidéo.

La section suivante trace l'état des lieux sur la disponibilité des corpus de LS, les deux sections suivantes sur la reconnaissance puis la génération, et la dernière section sur les applications existantes ou à venir [DALLE 2007].

CORPUS : DES LANGUES PEU DOTÉES

Les LS sont des langues peu dotées. En effet, elles ne disposent pas suffisamment, voire pas du tout pour certaines d'entre elles, des ressources dont peuvent disposer les langues, c'est-à-dire : un système d'écriture, des ouvrages de référence décrivant le fonctionnement de la langue (grammaires, dictionnaires), des ouvrages de diffusion massive (presse, œuvres culturelles), des ouvrages techniques et d'apprentissage (publications techniques, scientifiques, didactiques), des supports de communication du quotidien (courriers, modes d'emploi...), ainsi que des applications informatiques dans cette langue. De même, les corpus, qui constituent la seule façon de fixer et de garder une trace pérenne des productions de langue des signes d'un point de vue patrimonial, sont rares et de petite taille.

Si certaines langues des signes disposent d'ouvrage de références, d'œuvres culturelles sur des supports tels que le DVD et si quelques émissions télévisées sont parfois interprétées en LS sous la forme d'un médaillon incrusté dans l'image, cela reste très limité, et la plupart ne disposent même pas de ce peu de ressources. La recherche pâtit de cette situation ;

les ouvrages détaillant le fonctionnement de ces langues sont rares et très partiels, les corpus sont rares eux aussi et de petite taille, les systèmes de TALS existants ne sont que des prototypes de laboratoires, et sont rarement généralisables ou même réutilisables.

Les progrès technologiques pour la capture, le stockage et la manipulation des vidéos ont permis récemment (années 2000) d'initier plusieurs projets de création de corpus de LS dans différents pays et même des corpus multilingues (figure 1). Depuis peu, des ateliers, associés à des conférences ou financés sur des projets nationaux, permettent à la communauté scientifique d'échanger leurs expériences sur la création, l'annotation, l'analyse et l'archivage de corpus vidéo⁶. Certaines recommandations fournies par les chercheurs les plus aguerris commencent à être reprises dans différents projets [JOHNSTON 2008]. Il est cependant trop tôt pour définir des normes et des standards.

Figure 1: Extrait du corpus de LSF du projet européen Dicta-Sign⁷.



Le tableau ci-dessous liste quelques corpus de LS existants ou en cours de réalisation. Il en existe bien d'autres, mais qui atteignent rarement une taille de plus de quelques dizaines d'heures ou de locuteurs. Ce tableau illustre le caractère récent dans la plupart des pays des financements de réalisation de corpus de grande taille. Ces financements sont de courte durée, ce qui nécessite de cumuler une succession de projets, à l'exception notable de l'Allemagne qui bénéficie d'un financement de quinze ans. Ce dernier corpus, qui sera le plus important au niveau mondial, ne

6 <http://www.ru.nl/slcn>; <http://www.sign-lang.uni-hamburg.de/lrec2008/programme.html>

7 <http://www.dictasign.eu>

comportera que quatre cents heures de vidéos, ce qui est très loin de la taille des corpus audio ou écrits disponibles à ce jour.

Pays	LS	Nom corpus	Taille	Statut
Australie	Auslan	<i>Auslan Corpus</i> Financements variés depuis 1990 http://www.auslan.org.au	300h 100 locuteurs	Achevé
Grande-Bretagne	BSL	<i>BSL Corpus</i> Financement sur 3 ans : 2008-2010 http://www.bsllcorpusproject.org	249 locuteurs	Achevé
France	LSF	<i>Corpus Creagest</i> Financement sur 5 ans : 2007-2011 http://www.umr7023.cnrs.fr/-Realisation-de-corpus-de-donnes-.html	130h 125 locuteurs	En cours
Allemagne	DGS	<i>DGS Corpus</i> Financement de 15 ans : 2009-2023 http://www.sign-lang.uni-hamburg.de/dgs-korpus	400h 300 locuteurs	En cours

La constitution de corpus de très grande taille, de nature variée (lexique, monologue, dialogue, discussion de groupe), archivés et accessibles, est un des verrous les plus importants, à la fois pour l'archivage patrimonial de ces langues peu dotées, mais aussi pour permettre les recherches sur le TALS qui nécessitent, comme pour toutes les langues, l'analyse de corpus de grande taille.

ANALYSE ET RECONNAISSANCE

Le but ultime de l'analyse de vidéo en LS [ONG 2005] est de comprendre automatiquement le sens de l'énoncé, en vue de traduire cet énoncé dans une autre langue ou en vue d'effectuer une action, par exemple une requête à une base de données ou la recherche d'une information dans un document en LS. Dans le cas général, cette tâche est encore hors de portée des programmes informatiques. On est donc amené à traiter des domaines limités (champ sémantique et lexical réduit), à imposer des contraintes au niveau de l'expression et de son contexte ou à viser des

programmes seulement semi-automatiques venant assister un opérateur humain. Cependant, des étapes intermédiaires de traitement, plus accessibles, permettent déjà de produire des applications.

D'où viennent ces difficultés et comment s'explique l'écart de performance avec les programmes de TAL ou de traitement de la parole?

Tout d'abord du contexte de ces recherches. Comme indiqué en introduction, les LS ayant reçu récemment le statut de langue, les recherches sur les LS sont récentes. Aussi, les modèles linguistiques sur lesquels pourraient s'appuyer les modélisations informatiques ne sont pas encore stabilisés. Il y a encore peu de chercheurs en analyse d'image dont les LS constituent le thème de recherche et non un simple cadre applicatif. La quasi-totalité de ces chercheurs ne comprendraient d'ailleurs pas les LS supports de leur étude. Enfin, on sait que les outils de reconnaissance de la parole (langue vocale) ont fortement progressé lorsqu'ils ont pu intégrer des approches statistiques apprises sur de gros corpus. Il n'est pas possible, actuellement, de suivre la même approche dans le cas des LS, par manque de données [COOPER 2009].

L'autre difficulté est liée à la nature même du signal vidéo, très complexe à analyser. Elle provient de la mise en jeu de différents composants corporels, intervenant en parallèle, dont l'analyse doit s'effectuer à des échelles spatiales et temporelles très différentes (par exemple estimer à la fois un changement furtif de la direction du regard et une répétition de balancements du corps). De plus, la vidéo résulte de la projection sur un plan de postures et de mouvements 3D, ce qui entraîne une perte dans l'information tridimensionnelle et introduit de nombreuses occultations. Ce caractère spatial et multicomposants rend très aléatoire l'utilisation, pour la LS, des outils d'analyse et de reconnaissance mis au point pour des langues vocales, linéaires et mono source [DALLE 2006].

On présente généralement la chaîne de traitement en deux grandes étapes :

- l'**analyse**, qui consiste à détecter (identifier et suivre dans chaque image de la vidéo) les caractéristiques des composants corporels impliqués et à estimer leurs paramètres ;
- la **reconnaissance**, qui effectue une segmentation temporelle en unités et les identifie en les affectant à une classe. Ces unités linguistiques ont différents niveaux de granularité.

Cette reconnaissance est précédée d'une phase d'apprentissage où on a estimé les paramètres de ces classes à partir d'exemples et ou de connaissances a priori (règles de grammaire), dans le cadre d'un modèle linguistique donné de la LS.

Les premiers travaux ont porté sur la reconnaissance de l'alphabet dactylogique c'est-à-dire la réalisation par des configurations manuelles de l'alphabet de la langue écrite. Ce n'est donc pas réellement de la LS. Il s'agit de reconnaître des configurations des mains, isolées (lettres) ou enchaînées pour épeler un mot.

L'étape suivante a visé la reconnaissance de signes isolés, le plus souvent à partir de quatre paramètres manuels, la configuration, l'orientation, le mouvement et l'emplacement des mains par rapport au corps, le signe pouvant impliquer une seule main ou les deux. Certains travaux font état de reconnaissance de plusieurs centaines de signes avec des taux de reconnaissance de plus de 90 %. En réalité, les performances varient beaucoup avec la nature des données, selon que leur acquisition a été faite en 2D avec des caméras conventionnelles, en 3D avec des dispositifs de stéréovision ou enfin avec des systèmes multi-caméra de capture de mouvement avec marqueurs sur le corps et le visage. On n'utilise en général ces derniers dispositifs que pour apprendre les modèles (géométrie du corps, dynamique des mouvements, signatures des signes), la reconnaissance étant réalisée sur des vidéos 2D.

Le passage à l'analyse d'une production signée continue n'est pas direct : un énoncé en LS n'est pas une suite de signes isolés concaténés. Il met en jeu non seulement les signes manuels, mais aussi les postures et mouvements du tronc et de la tête et l'expression du visage. La direction du regard également joue un rôle grammatical important. Or tous ces éléments sont difficiles à détecter et à caractériser. Enfin le signeur⁸ utilise l'espace situé devant lui (espace de signation) comme support de son discours et pour le structurer. Des signes sont alors localisés dans cet espace et on observe de nombreuses opérations de pointage et de référencement. Là aussi, la perte de l'information 3D et la grande variabilité dans les manières possibles d'utiliser cet espace rendent difficiles sa caractérisation et la modélisation de son exploitation par le signeur [LENSEIGNE 2005]. En outre le traitement d'une production continue fait

8 La personne qui s'exprime en langage des signes.

apparaître les problèmes de transitions entre signes et de coarticulations [SEGOUAT 2010].

L'analyse d'énoncés en LS peut viser différents objectifs :

- **Annotation de corpus** : il s'agit d'assister l'annotateur en enrichissant le signal (reconstruction d'informations 3D), en réalisant automatiquement certaines mesures (dynamique des gestes, caractérisation des expressions du visage...) ou en détectant des événements particuliers (contacts mains, zones particulières du visage...).
- **Identification et mise en évidence de structures syntaxiques**, en particulier de l'exploitation de l'espace de signation (identifier les pointages et localiser leur cible...) et pour structurer les énoncés.
- **Reconnaissance de signes** : la recherche d'un signe dans un flux continu de LS met en œuvre des méthodes
 - de normalisation, pour s'affranchir des variations d'aspects et d'échelle des signeurs ;
 - d'alignement temporel, les signes pouvant être réalisés à des vitesses différentes ;
 - de caractérisation, pour extraire les propriétés intrinsèques du signe et non les variabilités introduites par chaque locuteur ;
 - de comparaison [ALON 2009].

Les méthodes les plus courantes sont basées sur les Modèles de Markov Cachés ou sur leurs variantes (couplés, parallèles), pour tenter de prendre en compte le parallélisme des signes, leur synchronisation et la spatialisation. Cependant ces méthodes, pour être performantes, devraient pouvoir s'appuyer sur une décomposition du signe en unités plus fines, de nature phonétique, dont la pertinence, la définition et la détection sont encore problématiques [THEODORAKIS 2010].

- **Compréhension de phrases** : les résultats dans ce domaine sont encore modestes [JUNG-BAE 2002], qu'il s'agisse de retrouver et d'expliquer les caractéristiques grammaticales d'une phrase ou d'essayer de la traduire dans une LV. L'ordre des signes dans un énoncé n'est pas le même que l'ordre des mots dans la langue vocale et il n'y a pas systématiquement une correspondance signe-mot. D'autre part, le signeur a la possibilité de choisir entre deux formes d'expression, une forme illustrative (« dire en montrant » en utilisant des structures exploitant l'iconicité)

et une forme non-illustrative (utilisant les signes standards). La forme illustrative faisant appel à l'expérience perceptivo-pratique, on peut se demander dans quelle mesure son interprétation par une machine est possible.

GÉNÉRATION ET ANIMATION

Les logiciels de génération de LS, couplés aux logiciels d'analyse présentés ci-dessus, permettent un usage entier, bidirectionnel, de la LS, que ce soit pour l'expression ou la compréhension en LS. Ils permettent aussi de rendre accessibles des informations disponibles à l'origine sous forme parlée, mais aussi sous forme écrite, étant donné que la grande majorité des adultes sourds ont du mal à maîtriser l'écrit. Les applications potentielles sont nombreuses : accessibilité du web, sous-titrage en LS, logiciels éducatifs en LS...

Ces logiciels procèdent en deux étapes : la génération de l'énoncé d'un point de vue linguistique, puis la génération du signal perçu, qui se présente sous la forme d'une animation d'un personnage virtuel, nommé « signeur virtuel ». Il n'existe pas encore de logiciels de génération de LS opérationnels comportant ces deux étapes, mais la recherche sur ce domaine est très active et en développement.

Pour la génération d'énoncé, on peut distinguer trois approches principales :

- **La génération d'énoncés par concaténation.** Cette méthode est utilisée dans le cas où les énoncés sont connus à l'avance, en nombre fini et comportent des parties variables, typiquement, les messages d'information ou d'alerte dans les lieux publics. C'est à ce jour l'approche la plus mature et elle est mise en œuvre en utilisation réelle en France (figure 2) pour les annonces en gares SNCF [BRAFFORT 2011]. Le système permet de générer des énoncés tels que « Le train n° xx en provenance de xx arrivera quai n° xx avec xx minutes de retard », les 'xx' représentant les parties variables de l'énoncé.

Figure 2: Diffusion de messages en LSF gare de l'Est à Paris.



- **La génération d'énoncés à partir de scripts.** Cette approche se base soit sur des outils d'éditions de type montage vidéo, soit sur des approches linguistiques [ELLIOTT 2008], [FILHOL 2010]. Cette dernière approche nécessite de disposer d'une base de données de descriptions lexicales et d'un ensemble de règles de grammaires décrivant le fonctionnement des LS. Ces aspects font encore actuellement l'objet de recherches fondamentales à long terme, les LS n'étant pas encore entièrement décrites.
- **La génération d'énoncés à partir de texte.** Il s'agit ici de méthodes de traduction automatique, souvent inspirées de celles utilisées pour les LV. Ces méthodes statistiques nécessitent de disposer de très gros corpus, qui ne sont pas disponibles à ce jour pour les LS. De plus, elles ont le défaut de ne pas du tout modéliser les spécificités des LS que sont l'iconicité, la spatialité et la multilinéarité. Une autre approche, plus récente [VENDRAME 2010], consiste à passer par une représentation pivot de nature sémantique qui semble plus proche de la manière dont s'organise le discours en LS [GUITTENY 2007].

Pour la génération des animations, on peut là aussi distinguer trois approches :

- **La génération d'animations par rotoscopie.** Cette approche s'appuie sur l'utilisation de vidéos comme modèles par un infographiste spécialisé, qui répercute à l'aide d'un logiciel spécialisé les postures et mouvements du locuteur filmé sur le personnage virtuel. Cette méthode est employée couramment dans la production de films d'animation 3D. Elle permet d'obtenir des animations de très grande qualité, indispensable pour une bonne compréhensibilité du contenu et pour une bonne acceptabilité par les locuteurs de LS. Son inconvénient est que

le résultat dépend de l'expérience et du talent de l'infographiste, et implique une certaine durée de production. Avec cette approche, il est possible de constituer une base de données d'animations que l'on peut ensuite concaténer et adapter. C'est cette méthode qui est employée actuellement dans le logiciel utilisé par la SNCF [BRAFFORT 2011].

- **La génération d'animations par capture de mouvement.** Cette méthode comme la précédente consiste à concaténer et adapter des animations prédéfinies, mais cette fois-ci à partir de bases de données de capture de mouvement [LU 2011]. Cette approche nécessite de la même manière la constitution préalable d'un corpus de LS, mais capté à l'aide d'un système à base de caméras infrarouges. L'avantage de cette méthode est qu'elle fournit directement des données en 3D, mais elle nécessite elle aussi une certaine durée de production car un « nettoyage » des données est indispensable avant leur utilisation. Elle nécessite de plus de disposer d'un système de capture de mouvement.
- **La génération d'animations automatique.** Il s'agit ici de générer les animations à partir d'une description symbolique. Cela relève de problématiques de recherche dans le domaine de l'informatique graphique et de la biologie (physiologie, modélisation du mouvement...). Quelques sites web incluent des logiciels construits sur cette approche, typiquement des sites de dictionnaires bilingues⁹. Les animations générées automatiquement sont encore à ce jour assez robotiques et manquent de réalisme en particulier sur la naturalité du mouvement et l'animation du visage. Dans ce dernier cas, cela est dû au manque de connaissance actuel sur le fonctionnement des LS concernant les composantes non manuelles, même si quelques études commencent à s'y intéresser [CHÉTELAT 2011].

Les avancées dans le domaine de la génération de LS sont incontestables et de plus en plus d'équipes de recherche s'y intéressent au niveau international. Pour permettre des avancées majeures, il est nécessaire de développer les études en amont sur l'analyse du fonctionnement des LS, par la constitution de gros corpus et le soutien aux études linguistiques et aux études pluridisciplinaires qui permettront d'intégrer les résultats de ces recherches dans leurs modèles et logiciels.

9 www2.cmp.uea.ac.uk/~jrwg/Dictionary094

CYBERESPACE : VERS DE NOUVELLES APPLICATIONS

Jusqu'à présent les dispositifs permettant aux sourds de vivre un bilinguisme, c'est-à-dire de pratiquer leur langue des signes, par exemple la LSF, dans un monde majoritairement entendant utilisant la langue vocale du pays, par exemple le français, ont surtout consisté dans des aides humaines et concernaient assez peu le cyberspace. Il s'agissait essentiellement de l'appel à des interprètes en présentiel.

Les nouvelles technologies et l'internet ont permis, avec les systèmes de visioconférence, d'étendre ces dispositifs : communication à distance entre personnes signantes et non signantes, via des interprètes dans des centres relais, ou directement en LS entre personnes signantes¹⁰. Cependant, dans tous les cas il ne s'agit que de communication entre personnes.

De nouveaux dispositifs [DALLE 2011] commencent à apparaître dans plusieurs domaines et portent maintenant sur l'accès à des contenus ou sur la LS elle-même [LEFEBVRE 2010B].

Concernant l'accès à des contenus, les premières applications ont porté sur l'accessibilité à des sites web existants. Elles consistaient à intégrer une aide en ligne sous forme de vidéos en LS traduisant une partie du document ou ajoutant des explications en LS (afin de palier les difficultés éprouvées par de nombreux sourds avec l'écrit). Cependant cette forme de complément en LS ne peut concerner que des pages évoluant rarement ; aussi, après un engouement initial, ce type de dispositifs ne s'est pas développé et le site web de la société WebSourd (figure 3), qui propose une traduction des brèves AFP en LS-vidéo tous les jours, reste une exception.

Les progrès en génération et en analyse de la LS ne se sont pas encore traduits par de nouveaux produits dans ce domaine, mais on peut déjà voir sur des maquettes expérimentales les apports que l'on peut en attendre :

- **Enrichissement de sites web :** des ACA (agents conversationnels animés) sont utilisés pour donner des indications de navigation ; des signeurs virtuels, visualisables à la demande, peuvent être utilisés pour apporter des compléments en LS¹¹, par exemple traduire la page FAQ d'un site. Ils présentent l'avantage, par rapport à la vidéo d'un signeur

¹⁰ <http://www.afils.fr/index.php/conseils> "Du bon usage des centres relais"

¹¹ <http://www.limsi.fr/Individu/jps/online/diva/geste/geste.main.htm>

réel, d'être anonymes, voire de respecter la charte graphique du site, et d'être modifiables, en particulier si le module de génération est alimenté par un module de traduction langue écrite / LS ;

- **Site wiki en LS :** Plus ambitieux encore, une des applications prototypes qui a été explorée durant le projet européen DictaSign est la réalisation d'un site wiki en LS. Il comporte une reconnaissance, au moins partielle, des informations données en LS par l'utilisateur via sa webcam et des messages générés et réalisés par un signeur virtuel ;
- **Sites bilingues :** les LS sont maintenant bien présente sur le Web, via des sites d'information bilingues de qualité¹². Outre leur caractère utile et culturel pour les personnes concernées, ils ont aussi pour effet de montrer les LS et de les rendre accessibles pour tous. Ils ont aussi un impact sur la langue elle-même : la langue utilisée est souvent plus académique que la langue des signes usuelle et ces sites ont un effet de diffusion des lexiques locaux et de propagation des néologismes ;

Figure 3: Utilisation intensive de la vidéo sur le site web de WebSourd.



- **Dictionnaires en LS :** le cyberspace est le lieu parfaitement adapté à ce type de document. Ils sont encore trop rares mais il y a plusieurs projets d'encyclopédies en cours de réalisation pour la LSF (Elix¹³, Ocelles [MOREAU 2010]) ou de glossaires thématiques (UVED¹⁴) ;

¹² <http://www.websourd.org>

¹³ <http://www.signesdesens.org/-e-learning-.html>

¹⁴ <http://www.irit.fr/GlossaireDD-LSF>

L'exploitation de ces contenus en LS comporte plusieurs volets [DALLE 2011]:

- Réalisation de ces contenus: il s'agit soit de documents en LS soit de documents bilingues. Il existe maintenant des outils faciles à utiliser [LEFEBVRE 2010B], permettant d'inclure un commentaire en LS dans un document vidéo (vidéo en LS, vidéo d'un texte ou d'une présentation);
 - Validation des signes utilisés: dans une langue en forte évolution et peu normalisée, cette question se pose avec acuité et les projets en cours prévoient tous des mécanismes d'identification de l'origine géographique des signes et de leur validation par la communauté sourde, en particulier pour les néologismes;
 - Présentation en LS: avant même d'exploiter un document en LS, il faut connaître son existence et pouvoir évaluer son intérêt. Les LS n'ayant pas d'écriture, outre l'usage du texte, des techniques particulières¹⁵ de présentation sous forme d'images, de photosignes (figure 4), de mini-vidéo permettent de présenter des tables des matières ou des index en LS;
 - Navigation dans des contenus en LS: le parcours d'un document en LS doit pouvoir se faire de manière non-linéaire: des expérimentations sont en cours¹⁶ pour reproduire dans une vidéo l'équivalent du lien cliquable dans un texte: il faut donc savoir comment faire apparaître ce lien puis comment l'utiliser. D'autre part les recherches en reconnaissance devraient permettre, à court terme, de faire des requêtes directement en LS, au moins sous forme de signes isolés.
- **Formation à distance en LSF:** au-delà de la mise en ligne de supports en LSF, l'enseignement en LSF à distance commence à se développer (en France: DAEU¹⁷ Nancy¹⁸, DU LSF Rouen¹⁹, DU IELS Toulouse²⁰). Si les cours sont encore le plus souvent sous forme de supports en LS (vidéo) à télécharger, le suivi des stagiaires utilise les ressources du web pour se faire directement en LSF (visioconférence, forum en LSF...) [TANAKA 2010];

¹⁵ http://www.usherbrooke.ca/liaison_vol41/n08/a_avaglyphe.html

¹⁶ <http://www.signlinkstudio.com/en/index.php>

¹⁷ Diplôme d'accès aux études universitaires – diplôme d'université.

¹⁸ <http://erudi.free.fr/index.php?page=daeulsf>

¹⁹ http://formations.univ-rouen.fr/LSA31_864/0/fiche___formation

²⁰ <http://www.irit.fr/iels>

- **Apprentissage de la LSF** : Les premiers sites en LSF étaient des sites de présentation de la langue et essentiellement des lexiques, donnant pour chaque entrée en français, le signe équivalent, les termes étant présentés par ordre alphabétique ou par thème. En LSF, il faut définir un ordre de présentation, le plus souvent de façon paramétrique, les signes étant classés, au premier niveau, suivant la configuration de la main (choix parmi une cinquantaine de configurations). On devrait bientôt pouvoir effectuer une requête en LSF via une webcam [LEFEBVRE 2010A]. Cependant ces sites sont d'un intérêt limité pour apprendre la LSF. Les sites d'enseignement de la LSF²¹ sont souvent des dépôts de cours enregistrés mis en ligne, les stagiaires pouvant envoyer leurs exercices sous forme de vidéo. Là aussi on peut attendre des développements intéressants pour l'auto-apprentissage, grâce aux outils de comparaison de signes et de génération par signeur virtuel [ARAN 2009];
- **Plateforme ressources pour les enseignants de ou en LSF** : la reconnaissance des LS et leur introduction dans l'enseignement (enseignement bilingue, option LSF au bac) a provoqué la création de sites ressources destinés aux enseignants (CNDP²², e-LSF²³) et utilisant largement la LS-vidéo mais encore peu les techniques avancées sur la LSF.

Figure 4: Exemples de photosignes « Développement » et « Durable » en LSF.



²¹ <http://www.signingsavvy.com>

²² <http://www.ressources-lsf.cndp.fr>

²³ <http://enseignement-lsf.com>

CYBERESPACE, UNE CHANCE POUR LES LS

La surdit  ne se voit pas et donc n'est pas prise en compte, sauf quand les sourds se mettent   s'exprimer en LS. Le cyberspace est une chance pour les LS :

- **Langues minoritaires, non localis es g ographiquement** : leur pratique est doublement p nalis e ; il est donc important de multiplier les lieux d'expression en LS, de recr er des espaces sociaux ;
- **Langues multim dia, dont la vid o constitue la forme «  crite »** : pour ces langues, le web et ses capacit s de diffusion d'images, d'animations et de vid os, ainsi que ses possibilit s d'interaction constitue un environnement privil gi  ;
- **Langues dont la composante grammaticale est assez universelle** : ce sont des langues facilitant les  changes internationaux.

Le cyberspace devrait permettre de procurer une  ducation en langue maternelle²⁴ aux jeunes sourds. La communication en langue des signes d s la naissance permet de restituer un environnement dans lequel les enfants sourds peuvent construire une vraie langue,   condition que la LS utilis e soit correcte. Seule la LS peut jouer ce r le de langue maternelle. Pour cela il faut qu'elle soit ou devienne aussi la langue des parents. La plupart des enfants n s sourds ont des parents entendants qui ne ma trisent pas (voire ne connaissent pas) la langue des signes ; ils doivent donc recevoir une formation pr coce et   domicile. Ces familles sont dispers es sur tout le territoire (pas de communaut  des familles d'enfants sourds) ; il est donc difficile de mettre en place des formations de groupe r guli res.

La formation   distance en LS et la diffusion de ressources et de situations de langage par internet apportent une r ponse   ce probl me. Des syst mes d'auto-apprentissage de la LS, int grant des m canismes de correction par reconnaissance ou de g n ration faciliteraient l'apprentissage de la LS par les familles et donc son acquisition pr coce par les enfants sourds, qui entreraient ainsi dans la vie avec une langue.

Les sourds pr sentent souvent un d ficit de formation qualifi e et de ma trise de la forme  crite des langues vocales de leur pays. L'existence de supports en LS, de sites ressources diffusant ces supports, d'encyclop die accessibles en LS et de formations en ligne pourraient combler

²⁴ Voir dans cet ouvrage : Marcel Diki-Kidiri, *Le cyberspace dans l' ducation en langue maternelle*.

ces manques. Si les techniques de fabrication de supports existent, il n'existe actuellement que peu de solutions opérationnelles sur l'accès à des contenus en LS (requêtes en LS, indexation, navigation en LS) ou sur leur génération (production de contenus par des signeurs virtuels). Or ces systèmes seraient très utiles pour la formation des personnes sourdes et leur développement culturel et donc leur intégration sociale.

Le développement de ces systèmes passe par une modélisation informatique plus complète et réaliste des LS, intégrant des modèles de variations, que ce soit pour reconnaître des énoncés spontanés ou pour générer des productions réalistes par signeur virtuel. Des campagnes de constitution de gros corpus de qualité devraient être lancées, afin d'accélérer l'élaboration de ces modèles, d'alimenter des programmes de recherche, en reconnaissance et en génération, et de permettre ainsi au TALS de se rapprocher des performances du TAL²⁵.

Les performances des réseaux, les capacités de stockage des serveurs du web et la puissance des ordinateurs personnels rendent maintenant possibles les échanges de vidéos fluides en langue des signes. La communication à distance entre personnes signantes est opérationnelle, de même que l'accès à des contenus pré-enregistrés en langue des signes. Par contre tout reste à faire pour permettre des interactions en langue des signes entre un usager et les ressources du cyberspace. L'absence d'écriture des langues des signes doit être compensée par des recherches sur des méthodes et des outils permettant de réaliser une réelle chaîne d'édition de documents en LS et donc d'alimenter le cyberspace avec des contenus en LS. L'accès à ces contenus doit être interactif en LS, ce qui implique là-aussi de développer des recherches sur la reconnaissance de requêtes en LS, pour que l'usager puisse prendre l'initiative, et sur la génération, par les systèmes, de réponses contextuelles donc produites par des signeurs virtuels. Dans ces conditions, le cyberspace sera une vraie ressource pour les sourds en leur permettant des utilisations dans leur langue et une vraie chance pour les LS en multipliant leur terrain d'expression et d'enrichissement.

25 Voir dans cet ouvrage: Joseph Mariani, *Les technologies de la langue en soutien au multilinguisme*.

BIBLIOGRAPHIE

- [ALON 2009] Alon J., Athitsos V., Yuan Q. & Sclaroff S. (2009). "A Unified Framework for Gesture Recognition and Spatiotemporal Gesture Segmentation", *IEEE Transactions of Pattern Analysis and Machine Intelligence (PAMI)* vol. 31 no 9 p. 1685-1699
- [ARAN 2009] Aran O., Ari I., Akarun L., Sankur B., Benoit A., Caplier A., Camp R, Carrillo AH., Fanard FX. (2009). "SignTutor: An Interactive System for Sign Language Tutoring", *IEEE Multimedia*, vol. 16 p.81-93.
- [BRAFFORT 2011] Braffort A., Bolot L. & Segouat J. (2011). "Virtual signer coarticulation in Octopus, a Sign Language generation platform". *International Gesture Workshop*, Athènes.
- [BRUGEILLE 2006] Brugeille JL., Dalle J. & Kellerhals MP (2006). "Une expérience d'utilisation de formes graphique dans la scolarité des enfants sourds : méthode de travail et premières observations", *Colloque Syntaxe, interprétation, lexique des langues signées*.
- [CHÉTELAT 2011] Chételat E. & Braffort A. (2011). "Investigation and analysis of non manual gestures involved in LSF: blinking". *International Gesture Workshop*, Athènes.
- [COOPER 2009] Cooper H., Bowden R. (2009) "Sign Language Recognition: Working with Limited Corpora", *Universal Access in HCI, Part III*, HCII 2009, LNCS 5616, pp. 472–481.
- [CUXAC 2000] Cuxac C. (2000). "La langue des signes française – les voies de l'icône". *Faits de langues* n°15-16 Ophrys.
- [DALLE 2005] Dalle P. (2005). "Histoire et philosophie du projet bilingue", *Nouvelle Revue de l'ALS*, Hors série "Enseigner et apprendre en LSF".
- [DALLE 2006] Dalle, P. (2006). "High level models for sign language analysis by a vision system", *2nd Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 5th edition of *Language Resources and Evaluation (LREC)*.
- [DALLE 2007] Dalle P, Braffort A., Collet C. (2007). "Accessibilité et langue des signes: modélisations, méthodes, application", *Conférence internationale sur l'accessibilité et les systèmes de suppléance aux personnes en situations de handicaps (ASSISTH 2007)*, Cépaduès, p.209-217.
- [DALLE 2011] Dalle P. (2011). "TIC au service de la LSF", colloque GERS Grandir et apprendre en LSF, *Revue Contact sourds entendants*, l'Harmattan.
- [ELLIOTT 2008] Elliott R., Glauer J.R.W., Kennaway J.R., Marshall I. & Safar E. (2008). "Linguistic modeling and language-processing technologies for Avatar-based sign language presentation". *Universal Access in the Information Society*, 6/4, Springer.
- [FILHOL 2010] Filhol M., Delorme M. & Braffort A. (2010). "Combining constraint-based models for Sign Language synthesis". *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 7th edition of *Language Resources and Evaluation Conference (LREC)*.
- [GUITTENY 2007] Guitteny P (2007). "Langue des signes et schémas", revue *Traitement Automatique des Langues (TAL)* Vol 48 2007. 3. "Modélisation et traitement des langues des signes". <http://www.atala.org/-Modelisation-et-traitement-des->
- [JOHNSTON 2008] Johnston T. (2008). "Corpus linguistics and signed languages: no lemmata, no corpus". *3rd Workshop on the Representation and Processing of Sign*

Languages: Construction and Exploitation of Sign Language Corpora, 6th edition of the Language Resources and Evaluation Conference (LREC).

[JUNG-BAE 2002] Jung-Bae K., Kwang-Hyun P, Won-Chul B., Zenn Bien Z. (2002). "Continuous Korean sign language recognition using gesture segmentation and Hidden Markov Model", FUZZ-IEEE'02. *IEEE Int Conf on Fuzzy Systems*, p. 1574-1579.

[LEFEBVRE 2010A] Lefebvre-Albaret F., Dalle P. (2010). "Requête vidéo dans une vidéo en langue des signes: Modélisation et comparaison de signes", RFIA.

[LEFEBVRE 2010B] Lefebvre-Albaret F., Dalle J., Piquet J., Dalle-Nazébi S., Gache P., Bacci A., Dalle P. (2010). "Analyse des langues des signes. Démarche de conception pluridisciplinaire d'outils d'analyse de discours en langues des signes", *Technique et science informatiques* (TSI) n° spécial L'informatique à l'interface de l'activité humaine et sociale Vol. 29 N° 8-9 pp.959-989

[LENSEIGNE 2005] Lenseigne B., Dalle P. (2005). "Using Signing Space as a Representation for Sign Language Processing", 6th *International Gesture Workshop*, Springer-Verlag, p. 25-3.

[LU 2011] Lu P, Huenerfauth M. 2011. "Collecting an American Sign Language Corpus through the Participation of Native Signers." *International Conference on Universal Access in Human-Computer Interaction (UAHCI)*.

[MOREAU 2010] Moreau C., Mascret B. (2010). "Organizing data in a multilingual observatory with written and signed languages". 4th *Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 7th edition of *Language Resources and Evaluation Conference (LREC)*.

[ONG 2005] Ong S.C., Ranganath S. (2005). "Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 6, pp. 873–891.

[SEGOUAT 2010] Segouat J. (2010). "Modélisation de la coarticulation en Langue des Signes Française pour la diffusion automatique d'informations en gare ferroviaire à l'aide d'un signeur virtuel", Thèse doctorat de l'université Paris-Sud 11, Orsay.

[SCHNEPP 2010] Schnepf J., Wolfe R. & McDonald J. C. (2010). "Synthetic Corpora: A Synergy of Linguistics and Computer Animation". 4th *Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 7th edition of *Language Resources and Evaluation (LREC)*.

[TANAKA 2010] Tanaka S., Matsusaka Y., Nakazono K. (2010). "Development of E-Learning Service of Computer Assisted Sign Language Learning: Online Version of CASLL" 4th *Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 7th edition of *Language Resources and Evaluation (LREC)*.

[THEODORAKIS 2010] Theodorakis S., Pitsikalis V., Maragos P. (2010). "Model-level data-driven sub-units for signs in vidéos of continuous Sign Language", *IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, p. 2262, 2265.

[VENDRAME 2010] Vendrame M. & Tiotto G. (2010). "ATLAS Project: Forecast in Italian Sign Language and Annotation of Corpora". 4th *Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, 7th edition of *Language Resources and Evaluation (LREC)*.

INTÉRÊT DES ARCHIVES ORALES POUR LES LANGUES MENACÉES

Le travail de la Fryske Academy (Académie Frison) et du Mercator European Research Centre sur le multilinguisme et l'apprentissage des langues est dédié à l'étude des langues minoritaires en Europe. La liste des langues courant un danger d'extinction a augmenté considérablement en Europe Centrale et Orientale, et plus encore au sein de la Fédération de Russie et à l'Est de la Sibérie. Cet article présente les projets scientifiques visant à l'utilisation de matériaux linguistiques issus d'archives ou de récolte de données sur le terrain.

Article original en anglais.
Traduit par Laurent Vannini.



TJEERD DE GRAAF est titulaire de doctorats en art, en linguistique, ainsi qu'en physique théorique. Il a été professeur associé en phonétique au département linguistique de l'Université de Groninge de 1975 à 2003.

Il effectue aujourd'hui ses recherches au sein de l'Académie Frison et du Mercator, Centre Européen de Recherche sur le Multilinguisme et l'Apprentissages des Langues.

TIJLDEED DE GRAAF

INTÉRÊT DES
ARCHIVES
ORALES POUR
LES LANGUES
MENACÉES

La **Fryske Akademy** (Académie Frisonne) et le Centre Européen de Recherche Mercator sur le Multilinguisme et l'Apprentissage des Langues, se consacrent tous deux à l'étude des langues minoritaires en Europe. L'engagement fondateur de l'Académie s'articule autour de l'histoire, de la littérature et de la culture apparentés au Frison occidental. Les locuteurs des deux langues voisines, le Frison oriental et le Frison septentrional parlées en Allemagne, sont moins nombreux et leurs langues sont répertoriées dans la liste des langues européennes en danger d'extinction. Une liste qui s'est considérablement allongée après l'entrée dans l'Union Européenne de nouveaux États d'Europe Centrale et Orientale. De même, plus loin à l'Est, au cœur de la Fédération de Russie et en Asie de l'Est, un grand nombre de langues sont identifiées comme étant en danger. Le présent travail décrit des projets existants ou à l'étude relatifs à quelques-unes de ces langues en danger de la Fédération de Russie et d'Asie de l'Est.

UNE PETITE HISTOIRE DES ARCHIVES SONORES.

Durant la seconde moitié du XIX^e siècle, Thomas Edison transforma radicalement les perspectives de la recherche en linguistique [DE GRAAF 1997, 2002C] en mettant au point en 1880 le phonographe, instrument capable d'enregistrer des sons. Pour la première fois, les êtres humains étaient en mesure de conserver et de réécouter des données acoustiques, vocales en particulier, et enfin de les reproduire pour être écoutées sur d'autres phonographes. Peu de temps après cette invention, des ethnologues, des folkloristes, des linguistes, des compositeurs ou même des amateurs passionnés commencèrent à utiliser cette nouvelle machine

pour collecter des informations sur les données orales et musicales de groupes culturels autour de chez eux ou à l'étranger.

Avant cette innovation, les linguistes sur le terrain devaient prendre des notes par écrit, un processus laborieux à la fois pour le chercheur et son informateur, puisque de nombreuses répétitions des mêmes énoncés oraux étaient nécessaires. Le phonographe a modifié la donne ; avec lui les linguistes pouvaient obtenir un enregistrement précis, objectif et instantané à partir d'une seule performance. Il était possible de saisir les nuances et subtilités de la langue orale, des reproductions pouvaient être jouées indéfiniment pour faciliter la transcription et l'analyse, tandis que les enregistrements originaux étaient préservés pour des usages futurs.

Pour de meilleurs résultats dans la reproduction des sons en provenance des vieux cylindres de cire, plusieurs appareils ont été conçus disposant de têtes de lecture légères pour réaliser l'extraction mécanique du signal. Afin de minimiser la dégradation des cylindres lors des lectures, et permettre la récupération de contenus à partir de cylindres cassés, des procédés optiques ont également été développés, autorisant des relectures sans contact physique et donc non destructives. Le premier d'entre eux fut inauguré par un laboratoire japonais de recherche [ASAKURA ET AL. 1986]. En 1988, j'ai été convié à travailler quelques mois avec ce groupe originaire de Sapporo (Japon) où j'ai pu tester la méthode développée localement sur plusieurs cylindres en cire, et bénéficier ainsi de l'expérience de mes collègues japonais.

De 1902 à 1905, à l'aide d'un phonographe, l'anthropologue polonais Bronislaw Pilsidski enregistra sur des cylindres en cire les conversations et les chansons des Aïnous vivant à Sakhaline et à Hokkaïdo, et ce afin d'étudier leur culture. Ces cylindres en cire ont été découverts en Pologne et transportés au Japon, où le groupe de recherche du professeur Asakura contribua à la reconstitution de ce matériel de grande valeur. Pendant mon séjour au Japon, le Professeur Kyoko Murasaki me permit de rencontrer les derniers locuteurs de l'Aïnou de Sakhaline, qui vivaient alors à Hokkaïdo [MURASAKI 2001] et me proposa de l'accompagner à Sakhaline pour y effectuer des recherches de terrain. Jusqu'en 1988, Sakhaline fut complètement isolée du reste du monde ; la perestroïka de Gorbatchev rendit ensuite possible la première expédition ethno-linguistique dans l'île, à laquelle je pris part en 1990 [DE GRAAF 1992]. Nous n'avons trouvé aucun vestige de la population Aïnou, mais visité plusieurs régions de

l'île où habite le peuple des Nivkhes. Les chapitres suivants de cet article témoigneront des projets relatifs à l'utilisation des archives sonores dans le cadre de l'étude de langues minoritaires.

QUELQUES PROJETS RELATIFS À L'UTILISATION D'ARCHIVES SONORES POUR L'ÉTUDE DE LANGUES EN DANGER D'EXTINCTION.

Notre groupe de recherche en phonétique et en ethnolinguistique a enquêté sur différents aspects des langues parlées dans la Fédération de Russie. Nous présenterons ici quelques uns des projets mis en œuvre pour l'étude des populations minoritaires de Russie et la description des langues en danger d'extinction. Dans le cadre de ces travaux, des données en provenance d'archives ont été utilisées et combinées à des résultats découlant de travaux de terrain contemporains dans plusieurs régions du Nord de la Russie, de la Sibérie, de l'Extrême-Est du pays et des territoires frontaliers entre la Russie et le Japon. Depuis 1992, ces projets ont été soutenus financièrement par la Netherlands Organisation for Scientific Research (NWO), par le Programme de soutien aux recherches internationales de l'Union Européenne (INTAS) et par l'entreprise Sakhalin Energy Investment Company Ltd. Nous avons collaboré avec des collègues en Russie et au Japon et une partie de notre travail est menée en parallèle de projets de recherche japonais.

Lorsque les premiers enregistrements ont été effectués, il devint évident qu'une infrastructure centrale serait nécessaire pour préserver les précieuses données recueillies. Au début du xx^e siècle, cela aboutit à l'établissement d'archives sonores, dont les plus anciennes en Europe sont localisées à Vienne, Berlin et Saint-Petersbourg. Les archives sonores de l'Académie Russe des Sciences situés dans l'enceinte du Musée de Littérature Russe (Pushkinsky Dom) à Saint-Petersbourg, contiennent environ 7 000 cylindres de cire gravés avec le phonographe d'Edison et 500 vieux disques de cire. De surcroît, un fond important de disques pour gramophone et l'une des collections les plus vastes d'enregistrements sur bande magnétique d'éléments du folklore de Russie témoignent de l'histoire de l'ethnographie russe et contiennent un large ensemble de matériels documentaires [DE GRAAF 2001, 2002A]. Une grande partie de ces

enregistrements constitue l'une des ressources principales utilisée dans nos projets réalisés en collaboration avec Saint-Pétersbourg.

Le premier de ces projets, sur l'utilisation de bases de données acoustiques et l'étude de l'évolution de la langue (1995-1998), fut soutenu financièrement par INTAS de l'Union Européenne. Nous avons reconstitué une partie des nombreux enregistrements archivés au Pushkinsky Dom afin de les rendre accessibles pour des recherches futures, ce qui est non seulement important pour des raisons culturelles et historiques, mais également pour aider à la description d'une langue et l'étude de potentielles preuves directes d'évolution d'une langue. Dans un second projet soutenu également par l'INTAS, *Les archives sonores de Saint-Pétersbourg sur le World Wide Web* (1998-2001), certains de ces enregistrements furent disposés sur l'internet et sont désormais accessibles sur un site dédié à de futures recherches [DE GRAAF 2004]. Le *Phonogrammarchiv* de l'Académie Autrichienne des Sciences fut à la fois partenaire et superviseur technique sur les deux projets.

Pour les réaliser, nous avons en premier lieu finalisé la reconstitution des matériels sonores archivés de la collection Zhirmunsky. Zhirmunsky fut un linguiste célèbre qui travailla à Saint-Pétersbourg/Leningrad au début du XX^e siècle. L'un de ses principaux centres d'intérêt était l'étude des dialectes allemands parlés en Russie. Entre 1927 et 1930, il enregistra de nombreux énoncés, des chansons de colons allemands en particulier, sur des disques en carton recouverts de cire qui furent ensuite transférés au Phonogrammarchiv de Vienne. Dans le cadre du projet INTAS, cette collection a été copiée sur bande magnétique et une partie du matériel est désormais stockée sur une base de données dédiée. Une étude spécifique eut pour objet la langue des Mennonites de Sibérie [DE GRAAF 2005].

Pour la mise en œuvre de notre troisième projet INTAS, la construction d'une base de données en texte intégral sur les langues balto-finnoises et dialectes russes du Nord-Ouest de la Russie (2000-2003), nous avons effectué un inventaire des langues minoritaires finno-ougriennes des environs de Saint-Pétersbourg, et des territoires au sud et au centre de la Carélie. Nous sommes ici devant une configuration linguistique singulière. Dans cette région plusieurs langues en danger comme le vepse, l'ingrien, le vote, le finnois-ingrien, le carélien et différents autres types de dialectes archaïques russes sont encore à ce jour parlées à proximité les unes des autres.

Les archives sonores de Saint-Pétersbourg contiennent également des données importantes sur le Yiddish, langue des juifs d'Europe de l'Est, qui au début du xx^e siècle était parlé par des millions de locuteurs dans l'empire Russe. Nous avons trouvé dans ces archives un manuscrit non publié, *Ballade dans le folklore juif*, qui était composé de matériel en yiddish enregistré sur des cylindres de cire. En collaboration avec des spécialistes de Saint-Pétersbourg nous avons exploré plus avant les données acoustiques présentes dans les archives sonores et préparé l'édition du livre. Une initiative qui s'inscrit dans le projet *Les voix du Shtetl, passé et présent de la langue yiddish en Russie* (1998-2001), pour lequel nous avons obtenu un soutien financier de la Netherlands Foundation for Scientific Research [DE GRAAF, KLEINER AND SVETOZAROVA 2004].

Le travail de terrain moderne et la reconstitution de données en provenance d'archives sonores fournissent d'importantes informations pour la préparation de descriptions d'une langue, sa grammaire, des dictionnaires et autres publications de littérature orale et écrite. Lors des expéditions menées dans les régions Nord de la Yakoutie, dans la région de l'Altaï et sur l'île de Sakhaline, nous avons étudié le processus de conversion linguistique et de disparition linguistique des populations autochtones de Russie, recueillant de nombreuses données de valeur.

LES LANGUES DE SAKHALINE

Notre première expédition internationale fut donc menée en 1990 sur l'île de Sakhaline, à la recherche des populations Aïnous, Nivkhes et Uilta afin de procéder à des enregistrements de leurs langues.

L'île de Sakhaline fait partie de la région de Sakhaline (*Sakhalinskaya Oblast'*), l'un des territoires les plus à l'Est de la Fédération de Russie. Sa superficie est de 87 100 kilomètres carrés et 980 kilomètres séparent le Nord de l'île de sa pointe Sud. Les îles Kuriles, un archipel de 1 200 kilomètres de long et composé de 36 îles, font également partie de ce territoire. La population originelle de Sakhaline était composée de tribus Paléosibériennes (ou Hyperboréennes) et Tougouses, avec en particulier les Nivkhes (anciennement les Guiliaks) et les Uilta (Oroks) au Nord et au Centre, et les Aïnous au Sud. Leurs populations étaient relativement faibles et durant la vague de colonisation par les Russes au Nord et par les Japonais au Sud, ils furent rapidement en infériorité numérique devant

ces nationalités plus puissantes. En raison de leur vie isolée loin du centre politique, elles furent cependant en mesure de maintenir leurs langues et cultures natives pendant une très longue période. Mais depuis le début du xx^e siècle le processus d'assimilation est devenu progressivement de plus en plus pressant.

Durant l'été 1990, je pris part à la première expédition visant à un travail linguistique de terrain sur l'île de Sakhaline, avec notamment comme objectif d'enquêter sur la situation linguistique et ethnographique des plus petites communautés nationales. L'idée était de trouver les derniers représentants de la population Aïnou ainsi que les membres d'autres groupes minoritaires, en particulier les Nivkhes et les Uilta. Malheureusement, nous ne pûmes trouver aucun membre de la communauté Aïnou, et la seule personne pouvant représenter la langue et la culture Aïnou de Sakhaline était probablement l'informateur que nous avons rencontré à Hokkaido, Asai Take san [DE GRAAF 1992], [MURASAKI 2001].

L'aïnou est la seule langue indigène du Japon qui soit en danger. Pour sa part, le nihkve fait partie des très nombreuses langues minoritaires de Russie. D'après les données démographiques disponibles, nous sommes parvenus à la conclusion qu'en 1989, les populations aborigènes du Nord représentaient une très petite minorité au sein de la population totale de Sakhaline: Le groupe ethnique Nihkve, le plus large d'entre les groupes minoritaires, ne représentait que 0,3 % des habitants insulaires [DE GRAAF 1992].

Parmi les nationalités à faible population de la Fédération de Russie, les peuples minoritaires du Nord jouent un rôle spécifique. Il existe près de trente groupes différents, tous habitant dans les zones les plus au nord du pays, en bordure de l'Océan Arctique, allant de la Scandinavie à la Mer de Béring et au Pacifique. Les populations du Nord furent tardivement placées véritablement sous la coupe Soviétique. Dans les années 30 le régime Soviétique tenta d'étendre son emprise sur ces populations et promut auprès d'elles la culture russe et l'apprentissage de la langue dominante. Un «Comité pour l'Assistance et la Protection des Petits Peuples du Nord» fut fondé en 1923 et un système d'écriture développé pour un grand nombre de langues minoritaires. L'alphabet latin fut le premier utilisé, puis fut remplacé dans les années trente par l'alphabet cyrillique.

La langue nivkhe est classée parmi les langues paléosibériennes. Elle est parlée par des tribus habitant aux abords du cours inférieur du fleuve Amour à l'Extrême-Est du continent asiatique, ainsi que dans les régions Nord et Centre de Sakhaline [GRUZDEVA 1998]. L'une de ses complexités linguistiques réside dans le fait que la langue connaît au moins deux dialectes : amour et sakhaline. Chaque groupe est relativement petit : 4 400 personnes en tout et pour tout ont la nationalité Nihkve, et 15 % d'entre eux parlent le nihkve. Une communauté très restreinte parle le dialecte de Poronaisk, ville du Sud de Sakhaline, pour laquelle il est très difficile de trouver des locuteurs. Après la guerre, certains d'entre eux émigrèrent de leur terre natale vers le Japon, où les Japonais et d'autres linguistes non soviétiques étudièrent leur langue.

Le premier recensement concernant l'ensemble des Russes fut organisé sous le régime du Tsar en 1897. Le nombre total de personnes répertoriées à Sakhaline comme appartenant au groupe ethnique Nivkhe, était alors de 1969. Tous identifièrent le nivkhe comme leur langue maternelle et la plupart étaient monolingues. Durant le recensement de 1926, le premier organisé par l'Union Soviétique, le nombre total de membres du groupe Nivkhe avait diminué, en raison du fait que les habitants de la partie sud de l'île, devenue japonaise, ne furent pas comptés. Tous pratiquement désignèrent le nivkhe comme leur langue première. Depuis lors, cependant, un déclin dans le pourcentage de locuteurs nivkhe a pu être observé même si le nombre de personnes de la communauté est resté stable sur l'île de Sakhaline (environ 2000). En 1989, plus de 80 % de la population Nivkhe ne parlait plus le nivkhe, et identifiaient le russe comme leur langue première.

La transition entre le nivkhe de Sakhaline et le russe peut s'expliquer de plusieurs façons. L'un des facteurs les plus importants se trouve dans la proximité grandissante entre la population Nivkhe et les autres habitants de l'île, dont une grande partie venait du continent, parlait russe et avait migré sur l'île pour y exploiter ses ressources naturelles (huile, charbon, bois, poisson, caviar). Avant cela, les Nivkhes pratiquaient la pêche et la chasse dans leurs villages isolés, puis ils eurent des contacts de plus en plus fréquents avec les immigrants, qui initièrent une politique active d'éducation et de mise sous influence des habitants autochtones des régions les plus à l'Est de la Fédération de Russie.

Une évolution en faveur des langues et cultures des minorités de la Fédération de Russie, en particulier le nivkhe [DE GRAAF AND SHIRAISHI 2004], est à l'œuvre depuis peu. Des expériences sont menées pour redonner vie à la langue nivkhe, avec par exemple la mise en place de classes de nivkhe dans plusieurs écoles. En 1980, le Ministère de l'Éducation de la Fédération de Russie initia un programme destiné aux écoles primaires et secondaires, pour lesquelles livres de classe et dictionnaires furent imprimés en nivkhe. Des instructions spécifiques furent données aux enseignants d'ascendance Nivkhe à propos de l'éducation dans leur langue des enfants Nivkhe. Ce programme éducatif fut introduit dans les internats pour enfants Nivkhe des minorités ethniques résidant à Nogliki, Chir-Unvd et Nekrasovka. Nous avons pu visiter ces écoles et appréhender les méthodes d'enseignement utilisées dans l'éducation primaire.

Durant notre expédition sur le terrain à Sakhaline, un abondant matériel linguistique concernant les langues des communautés minoritaires fut collecté. La majorité des sujets participants étaient des personnes âgées extrêmement motivées à l'idée d'utiliser leur langue, comme par exemple des membres d'un groupe folkloriques. Pratiquement aucun des jeunes que nous avons rencontrés n'avait de connaissance active de la langue, et elles ne communiquaient avec leurs parents qu'en russe. Les informateurs Nivkhe, que nous rencontrions au cours d'entretiens se montraient extrêmement convaincus de l'importance que revêtaient la conservation et le développement de leur propre culture, et ils exprimèrent le désir de participer en tant que membres d'un groupe des nations composant la Fédération de Russie. Ils s'accordaient sur le fait que la langue et la culture russe jouaient un rôle important dans leur vie, mais indiquaient vouloir sauvegarder leur langue et leur culture native par tous les moyens possibles.

DES VOIX VENUES DE LA TOUNDRA ET DE LA TAÏGA

D'importantes activités liées aux bases de données linguistiques de Saint-Pétersbourg s'appuient sur les enregistrements de dialectes russes et de langues minoritaires de Fédération de Russie, comme le nivkhe, les langues tongouses, le yakoute ou d'autres encore [DE GRAAF 2004]. L'un de nos objectifs est d'utiliser ces enregistrements afin de constituer une base de

données phonétique des langues de Russie. Dans le cadre du programme de recherche *Voix de la toundra et de la taïga*, qui débuta en 2002, nous avons combiné les données en provenance de vieux enregistrements sonores aux résultats de nos enquêtes de terrain afin de présenter un panorama descriptif complet des langues et cultures des groupes ethniques en Russie. Les langues et cultures de l'Arctique, en danger au sein de la Fédération de Russie, doivent être décrites rapidement avant qu'elles ne soient définitivement éteintes. Notre travail antérieur sur des technologies de reconstruction de vieux enregistrements sonores découverts dans les archives de Saint-Pétersbourg a rendu possible la comparaison de langues encore parlées dans les territoires de nos recherches à celles qui y étaient parlées il y a plus d'un siècle, ce qui a fourni un point de départ providentiel à notre étude. Les enregistrements sonores contenus dans les archives de Saint-Pétersbourg consistent en un ensemble de matériels conversationnels, de chansons folkloriques, de contes de fées, principalement dans des langues sibériennes [BURYKIN ET AL. 2005], [DE GRAAF 2004A].

Cartes des langues de Sibérie¹



1 Cartes des langues de Sibérie, avec la gracieuse permission de l'Institut Max Planck, Leipzig. © MPI for Evolutionary Anthropology.

Dans la mise en œuvre de ces projets, les techniques développées précédemment ont été appliquées à quelques-unes des langues et cultures minoritaires de Russie en voie de disparition, comme le nivkhe et l'uilta sur Sakhaline, et les langues tOUNGouses et youkaguire en Yakoutie. Notre but est la mise en place d'une phonothèque et d'une vidéothèque d'enregistrements ayant trait aux histoires, folklore, traditions orales et chants des peuples de Sakhaline et de Yakoutie. Les enregistrements sonores composant les déjà archives de Sakhaline et de Yakoutie seront complétés par les résultats de ces nouvelles enquêtes de terrain. Les données obtenues seront ajoutées aux matériels archivés à l'heure actuelle à Saint-Pétersbourg, et seront rendues en partie accessible sur l'internet et sur CD-Rom.

Ce projet de recherche et la documentation relative sont mis en œuvre en coopération étroite avec des chercheurs dans des centres locaux comme celui de Yuzhno-Sakhalinsk qui participe en archivant les enregistrements sonores et les enquêtes de terrain. Des spécialistes en provenance de Saint-Pétersbourg et des Pays-bas les rencontrent régulièrement pour implanter de nouveaux centres dédiés à l'étude et l'enseignement des langues locales ou de thématiques associées. Dans ce but, nous avons organisé un séminaire spécialement pour les enseignants du nivkhe à Yuzhno-Sakhalinsk en octobre 2003.

Séquences spontanées de conversation et lectures de textes préparés sont collectés dans le cadre d'analyses ethnolinguistiques, mais également anthropologiques, folkloristes et ethnomusicologiques. Ces données sont filmées et analysées, et peuvent illustrer l'art du conte et l'utilisation de la langue. Les textes décrits plus hauts seront publiés dans des revues et livres scientifiques, accompagnés d'illustrations audiovisuelles sur CD-Rom ou sur l'internet. Les matériels seront dès lors accessibles pour de plus amples analyses à des chercheurs travaillant dans le champ de la phonétique, de la linguistique, de l'anthropologie, de l'histoire de l'ethnomusicologie et du folklore.

À l'aide d'un manuel de conversation scolaire pour enfants de la communauté Nivkhe [TAKSAMI ET AL. 1982], nous avons enregistré un locuteur natif pendant notre séjour de terrain en 1990. Les textes et illustrations du livre sont désormais consultables sur internet, ainsi que les données acoustiques. Les phonèmes isolés sont disposés dans un tableau spécifique; en sélectionnant l'un d'entre eux, l'étudiant peut écouter différents

extraits de conversations. Ainsi, les étudiants seront capables d'apprendre la distinction entre les divers phonèmes séparés du nivkhe (par exemple les quatre sons en K), qui sont des variantes (allophones) d'un seul phonème en langue russe. L'un de nos étudiants chercheurs et ses collègues Nivkhes publièrent une série de livres d'histoires, de chansons et de conversations propres à la communauté, dans lesquels pour la première fois les textes correspondants sont enregistrés sur un CD. La série intitulée *Matériaux sonores pour la langue nivkhe I – III* [SHIRAISHI, LOK 2002, 2003, 2004] fut publiée dans le cadre d'un programme japonais sur les langues en danger de la zone pacifique (ELPR) et trouva également sa place dans le programme de recherche *Voix de la toundra et de la taïga*. Ce matériel unique est utilisé non seulement par des linguistes, mais aussi par la communauté linguistique elle-même, où elle sert de ressource pédagogique. En 2006, Hidetoshi Shiraish a rédigé sa thèse sur le sujet, *Aspects of Nivkh Phonology*, et l'a soutenue en septembre 2006 à l'Université de Groninge [SHIRAISHI 2006].

ARCHIVES EN DANGER

Durant l'été 2005, nous avons présenté à la Netherlands Organisation for Scientific Research, la nwo, le projet de recherche *Voix de la toundra et de la taïga*, et publié un catalogue des enregistrements existants d'histoires, de folklore et chants et traditions orales des peuples de Sibérie [BURYKIN ET AL. 2005]. Ces ressources ont ainsi été rendues accessibles pour d'autres analyses par des chercheurs travaillant dans le champs de la phonétique, de la linguistique, de l'anthropologie, de l'histoire, de l'ethnomusicologie et du folklore. L'information est également d'une grande valeur pour le développement de méthodes pédagogiques destinées aux représentants des groupes ethniques concernés et pour la conservation et la revitalisation de leurs langues et de leurs cultures.

Actuellement, un grand nombre des vieux enregistrements reste inaccessible dans des archives et lieux privés où la qualité de conservation n'est aucunement garantie. Lors d'un projet mené de septembre 2006 à septembre 2008, et aidé financièrement par un programme de la British Library pour les langues en danger, nous avons rendu public une partie de ces enregistrements et les avons ajoutés à la base de données développée à Saint-Pétersbourg.

Notre partenaire sur ce projet a une nouvelle fois été le *Phonogrammarchiv* de l'Académie autrichienne des Sciences. L'objectif était de réenregistrer le matériel en utilisant des technologies contemporaines [SCHÜLLER 2005], et de le stocker dans un endroit sûr avec les métadonnées obtenues à partir des fiches et références correspondantes. L'infrastructure de stockage mise à disposition par le projet pourrait moderniser les activités potentielles d'archivage dans la Fédération de Russie et inciter à l'adoption de standards technologiques plus modernes.

Les originaux sur bandes magnétiques et cassettes audio furent copiés sur des disques durs. Au sein d'une collection de plus de 111 heures (et 218 Go) de matériel sonore numérisé, les langues suivantes sont représentées : l'azerbaïdjan, le baloutche, le djaghataï, le chatror, Le dari, l'énetse, le kati, le kerek, le mendzon, le nénétsse, le nganassane, le parachi, le pashai, le pachtoune, le russe, le shughni, le tadjike, l'oudihe, le vaygali et le wakhi (vakhan). Les données de ces archives sonores numériques fournissent des informations sur le développement historique des langues concernées et peuvent être utilisées pour la description linguistique, l'étude du folklore et l'ethnomusicologie d'un grand nombre de langues minoritaires en danger de Russie.

Il existe des collections similaires dans d'autres régions de la Russie, non seulement dans des instituts publics, mais également entre les mains de collectionneurs privés où elles manquent souvent d'une protection adéquate. C'est le cas par exemple des collections privées sur le nivkhe, accessibles à Yuzhno-Sakhalinsk, à Vladivostok, à Londres ou ailleurs. Pour la majorité de ces ressources, on peut estimer que la qualité de conservation est inférieure aux normes standards. Dans le sillage de notre collaboration de longue date avec les chercheurs de Sakhaline, nous envisageons de créer à Yuzhno-Sakhalinsk des infrastructures de stockage de matériels sonores liés aux langues aborigènes de l'île. Les collections les plus importantes sont celles citées plus haut et concernent la langue nivkhe, mais nous devons y ajouter les ressources sonores sur les Aïnou et les Uilta de Sakhaline. Nous ne connaissons pas la taille exacte de certaines de ces collections privées, mais pour d'autres une estimation préliminaire est nécessaire. Dans le cadre de ce projet et de projets futurs, nous espérons obtenir accès à ces collections, les copier sur des supports audio plus modernes, rendre disponible un catalogue et publier une partie de ce matériel avec les enregistrements du même type

provenant de Saint-Pétersbourg. Sur Sakhaline et dans d'autres régions de Russie, les chercheurs locaux seront impliqués dans la préparation de ces projets avec le soutien de collègues de Saint-Pétersbourg, d'Autriche, des Pays-Bas et du Japon.

QUELQUES REMARQUES POUR CONCLURE

Nous avons décrit ci-dessus les efforts conjoints de chercheurs de Russie et des Pays-Bas dans l'analyse de données en provenance d'archives sonores et dans l'utilisation de techniques modernes d'enquêtes de terrain pour l'étude de langues en danger telles que le nivkhe, le nénése ou le youkaguire. Les résultats obtenus sont des descriptions de langues, des grammaires, des dictionnaires et des collections publiées de littérature orale et écrite sur et dans ces langues. Lors de séminaires, l'usage de ces matériels d'enseignement et d'apprentissage transcrits sur des équipements modernes de technologie de l'information peut être transmis à des enseignants locaux. L'enseignement des anciennes langues premières peut servir aux plus jeunes membres des communautés qui n'ont pas appris leur langue native en famille. En Russie, des méthodes spécifiques d'enseignement des anciennes langues maternelles comme langues étrangères doivent être appliquées. Une partie des bases de données acoustiques utilisées pour certains des projets sont disponibles en ligne et donnent la possibilité d'échanger de l'information sur ces langues avec des institutions à l'autre bout du monde.

Au niveau de la communauté locale, et ce depuis plusieurs décennies, un grand nombre de personnes travaillent au développement de programmes d'éducation linguistique, généralement avec des ressources techniques extrêmement réduites. Contrairement à des enseignants de langues dominantes ailleurs dans le monde, ils manquent non seulement de formation officielle dans l'enseignement d'une langue, élément désormais couramment requis par les gouvernements locaux, mais également de programmes propres à la langue et, plus crucial encore peut-être, de descriptions utilisables des bases de la langue. Le Centre Européen de Recherche Mercator a l'intention de jouer un rôle dans la coordination de ces diverses activités. Il sera important d'échanger des idées avec des instituts similaires dans d'autres parties du monde. Ensemble, nous devrions être capables de développer une stratégie efficace et viable pour soutenir les langues en danger de disparition dans le monde.

BIBLIOGRAPHIE

[**ASAKURA ET AL. 1986**] Reproduction of sound from old wax phonograph cylinders using the laser-beam reflection method. *Applied Optics*, vol.25, no. 5, pp.597 – 604.

[**AUSTIN, PETER K. (ED.) 2002 AND 2004**] Language Documentation and Description. *The Hans Rausing Endangered Languages Project*, SOAS, London. (Vols. 1 and 2).

[**BURYKIN ET AL. 2005**] Burykin, A., A. Girfanova, A. Kastrov, I. Marchenko and N. Svetozarova 2005. *Kollektsii narodov severa v fonogrammarxive pushkinskogo doma*. [Collections on the peoples of the North in the phonogram archive of the Pushkinski Dom]. Faculty of Philology, University of St.Petersburg.

[**DENISOV VICTOR. 2008**] Zapisi udmurtskogo jazyka i folkloru v Fonogrammarkhive Istituta russkoi literatury (Pushkinskij Dom). In: *Rossija i Udmurtija: istorija I sovremennost'* [Recordings of the Udmurt language and folklore in the Phonogram Archive of the Institute of the Russian literature (Pushkinskij Dom). In: *Russia and Udmurtia: History and Present*]. Izhevsk, 879-884.

[**DE GRAAF 1992**] The languages of Sakhalin. Small languages and small language communities: news, notes, and comments. *International Journal of the Sociology of Languages* 94, pp. 185-200.

[**DE GRAAF 1997**] The reconstruction of acoustic data and the study of language minorities in Russia. In: *Language Minorities and Minority Languages*. Gdansk: Wydawnicstwo Uniwersytetu Gdanskiego, pp 131-143.

[**DE GRAAF 2001**] Data on the languages of Russia from historical documents, sound archives and fieldwork expeditions In: Murasaki, K. (red.) *Recording and Restoration of Minority Languages, Sakhalin Ainu and Nivkh*, ELPR report, Japan, pp. 13 – 37.

[**DE GRAAF 2002**] The Use of Acoustic Databases and Fieldwork for the Study of the Endangered Languages of Russia. Conference Handbook on Endangered Languages, Kyoto. *Proceedings of the Kyoto ELPR Conference*, pp. 57-79.

[**DE GRAAF 2002**] Phonetic Aspects of the Frisian Language and the Use of Sound Archives. In: *Problemy i metody eksperimental'no-foneticheskikh issledovanij*. St.Petersburg, pp. 52-57.

[**DE GRAAF 2002**] The Use of Sound Archives in the Study of Endangered Languages. In: *Music Archiving in the World*, Papers Presented at the Conference on the Occasion of the 100th Anniversary of the Berlin Phonogramm-Archiv, Berlin, pp. 101-107.

[**DE GRAAF 2004**] Voices from Tundra and Taiga: Endangered Languages of Russia on the Internet. In: Sakiyama, O and Endo, F (eds.) *Lectures on Endangered Languages: 5 - Endangered Languages of the Pacific Rim* C005, Suita, Osaka, pp. 143-169.

[**DE GRAAF 2004**] The Status of Endangered Languages in the Border Areas of Japan and Russia. In: Argenter, J A and McKenna Brown, R (eds.) *On the Margins of Nations: Endangered Languages and Linguistic Rights*. Proceedings of the Eighth Conference of the Foundation for Endangered Languages, Barcelona, pp. 153-159.

[**DE GRAAF 2004**] With Kleiner, Yu. And Svetozarova, N. Yiddish in St.Petersburg: *The Last Sounds of a Language*. Proceedings of the Conference "Klezmer, Klassik, jiddisches Lied. Jüdische Musik-Kultur in Osteuropa." Wiesbaden, Harrassowitz Verlag, pp. 205 - 221.

[DE GRAAF 2005] Dutch in the Steppe? The Plautdiitsch Language of the Siberian Mennonites and their Relation with the Netherlands, Germany and Russia. In: Crawhall, N. and Ostler, N. (eds.): *Creating Outsiders. Endangered Languages, Migration and Marginalisation*. Proceedings of the IXth Conference of the Foundation for Endangered Languages, Stellenbosch, 18-20 November 2005, pp. 32-31.

[DE GRAAF 2004] With H. Shiraiishi. Capacity Building for some Endangered Languages of Russia: Voices from Tundra and Taiga. In: *Language Documentation and Description, Volume 2*, The Hans Rausing Endangered Languages Project, School of Oriental and African Studies, London, pp. 15-26.

[DE GRAAF 2008] With V. Denisov. Sokhraneniye zvukovogo nasledija narodov Udmurtskoi Respubliki: opyt veduschikh zvukovykh arkhivov mira. *Rossija i Udmurtija: istorija i sovremennost'* [Preservation of the sound heritage of the peoples of the Udmurt Republic: experience of the World's leading archives]. In: *Russia and Udmurtia: History and Present*. Izhevsk, 866-878.]

[GRUZDEVA 1998] *Nivkh*. München: Lincom Europa.

[IASA-TC 03. 2004] *The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy*. International Association of Sound and Audiovisual Archives (IASA).

[IASA-TC 04. 2007] *Guidelines on the Production and Preservation of Digital Objects*. International Association of Sound and Audiovisual Archives (IASA)

[MAGID, S. 1936] Spisok Sobranii Fonogramarkhiva Folklornoi sektsii IAE Akademii Nauk SSSR [List of the Collections in the Phonogram Archive of the Folklore Section of the Institute for Anthropology and Ethnographics, Academy of Sciences of the USSR]. *Sovetskii Folklor*, №4-5, pp. 415-428.

[HINTON, L. 1994] *Flutes of Fire. Essays on California Indian languages*. Berkeley, California: Heyday Books.

[MATSUMURA, KAZUTO (ED) 1998] *Studies in Endangered Languages*. Papers from the international symposium on endangered languages, Tokyo, November 18-20, 1995. Tokyo: Hituzi Syobo.

[MURASAKI, K. 2001] *Tuytah: Asai Take kojutsu, Karafuto Ainu no mukashi banashi* (Old stories of the Sakhalin Ainu). Tokyo: Sofukan.

[SCHÜLLER, D. (ED.) 2005] *The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy*. IASA Technical Committee – Standards, Recommended Practices and Strategies, IASA-TC 03.

[SHIRAIISHI, H. AND G. LOK. 2002, 2003] *Sound Materials of the Nivkh Language 1 and 2*. Kyoto, ELPR Publications A2-15 and 36.

———. 2004. With G. Lok. *Sound Materials of the Nivkh Language 3*. Publication of the International NWO project “Voices from Tundra and Taiga”, University of Groningen.

———. 2006. *Topics in Nivkh Phonology*. Dissertation at Groningen University, September 2006.

[TAKSAMI, CH.M., PUKHTA, M.N., VINGUN, A.M. 1982] *Nivkhgu bukvar'*. Leningrad, Prosveshchenie.

[TAMURA, S. 2000] *The Ainu Language*. Tokyo: Sanseido.

[WITSEN, N. 1705] *Noord en Oost Tartarye, ofte bondig ontwerp van eenige dier landen en volken, welke voormaels bekend zijn geweest*. Amsterdam, Halma.

POLITIQUES LINGUISTIQUES POUR CONTRER LA MARGINALISATION DES LANGUES

L'urbanisation et la globalisation promeuvent l'unification des cultures ethniques et rejettent fortement une grande majorité d'entre elles dans les marges. Les connaissances, ainsi que l'expérience historique et culturelle de ces cultures et de leurs langues s'amenuisent. Une culture singulière disparaît lorsque sa langue meurt. Aussi le développement de technologies de la communication est porteur d'espoir et d'opportunités. Quelles actions peuvent être menées pour stopper ou même seulement ralentir la marginalisation des langues, en améliorer la vitalité, la représentation et l'utilisation ? Qui peut le faire et surtout qui en a le devoir ?

Article original en anglais.
Traduit par Laurent Vannini.



EVGENY KUZMIN est président du Conseil intergouvernemental du Programme Information pour tous au sein de l'Unesco. Il a dirigé le Département des bibliothèques du Ministère de la Culture de Russie entre 1992 et 2005 et participé à l'élaboration de la politique culturelle nationale russe, notamment des stratégies de construction de la société de l'information. Il est à l'origine de plusieurs projets portant sur la diversité linguistique dans le cyberspace, dont deux conférences internationales.

EVGENY KUZMIN

DES LANGUES
MARGINALISÉES
POUR CONTRE
LINGUISTIQUES
POLITIQUES

Un grand nombre de nations du monde contemporain n'ont ni État, ni souveraineté. Leurs langues ne sont pas les langues officielles de leur pays de résidence puisqu'un grand nombre de pays sont multiethniques et multilingues. Dans le meilleur des cas, même lorsque le gouvernement et les groupes ethniques dominants accordent une attention poussée aux minorités ethniques et linguistiques, la plupart des langues sont marginalisées et ce à des degrés variés. Elles se développent ou déclinent dans l'ombre de la langue majoritaire et dominante du pays, celle qui est utilisée dans toutes les sphères de la société – politique, économique, culturelle, scientifique, etc.

La mondialisation, les migrations et une urbanisation effrénée donnent aux membres d'un grand nombre de minorités ethniques le sentiment que leur langue manque de prestige. L'apprentissage des langues officielles ou internationales concentre quasiment toute l'attention et les efforts de recherche.

Aucune langue ne peut se développer en dehors de la culture de l'ethnos qui l'a créée et la parle. Parallèlement, l'urbanisation et la mondialisation promeuvent l'unification des cultures ethniques autour de la culture majoritaire, les poussant ainsi plus loin encore dans la marge. Les connaissances, ainsi que l'expérience culturelle et historique véhiculées par ces cultures et ces langues s'évanouissent progressivement, et leur futur à la fois comme langue et culture est menacé. Marginalisation linguistique et marginalisation culturelle sont deux aspects interconnectés d'un processus à visage multiple. Une culture unique disparaît à la mort de la langue qui la porte⁴.

1 Le terme « culture » est utilisé ici dans son sens le plus large, soit l'ensemble des caractéristiques culturelles, matérielles, intellectuelles et émotionnelles les plus saillantes d'un groupe social ou d'une communauté en particulier. La culture englobe non seulement les

Ces questions sont d'actualité principalement au sein de pays où deux langues au moins sont pratiquées. Mais la situation n'est pas sans espoir. Le développement des technologies de l'information et de la communication (TIC) apporte un regain d'optimisme.

Que pouvons-nous faire pour stopper, ou pour le moins ralentir la marginalisation des langues, et renforcer la vivacité, la représentation et l'usage des langues en voie de disparition ? Qui peut intervenir, et pour qui est-ce un devoir d'intervenir ? Prenons la Russie comme exemple – l'un des pays constitué de la plus grande diversité ethnique, linguistique et religieuse du monde – et analysons la manière dont elle se pose la question et les réponses qu'elle y apporte.

LE PAYSAGE LINGUISTIQUE RUSSE

Selon une estimation de 2009, la Fédération de Russie compte environ 141,9 millions d'habitants. Les membres de l'ethnie Russe y contribuent pour près de 80 % (114 millions). Cent quatre-vingts langues sont parlées parmi les autres entités ethniques. Un peu plus d'une centaine de ces langues appartiennent aux entités ethniques indigènes constituées historiquement à l'intérieur des frontières actuelles de la Russie, ou y vivant depuis des siècles. Cependant, plus de 127 millions de personnes considèrent la langue officielle du pays – le russe – comme leur langue première. Beaucoup de membres des populations indigènes maîtrisent mieux le russe que leur propre langue maternelle ; certains la parlant même mieux que des membres de l'ethnie Russe.

Les langues les plus répandues en Russie sont le tatar (5,35 millions de locuteurs), le bachkir (1,38 millions), le tchétychène et le tchouvache (1,33 millions chacune). Il y a également neuf autres langues pour lesquelles le nombre de locuteurs est compris entre 400 000 et 1 million. L'avar, le kabarde-tcherkesse, le darguine, l'ossète, l'oudmourte, le koumyk, le yakoute, le mari, l'ingouche. Quinze autres langues indigènes sont parlées par 50 à 400 000 personnes.

Toutes les langues parlées en Russie, sauf le russe, sont des langues minoritaires et sont touchées par la marginalisation à des degrés divers. Plus d'un tiers des langues parlées en Russie sont en voie de disparition ou

arts et la littérature, mais également le mode de vie, les droits humains fondamentaux, le système de valeurs, l'éducation, les coutumes, les traditions et la philosophie.

sérieusement en danger. La situation est encore pire pour 39 langues des minorités ethniques fortes de moins de 50 000 locuteurs et qui sont principalement pratiquées par des populations indigènes de la Région du Nord, de la Sibérie et de l'Extrême-Orient russe. Bien que les autorités russes à tous les échelons prêtent une attention singulière à ces langues et cultures concernées, le risque de les voir s'éteindre ne doit pas être sous-évalué.

La constitution de la Fédération de Russie déclare toutes les langues de Russie comme des biens culturels communs. Presque toutes ces langues ont une écriture, que certaines ont développée assez récemment.

Au contraire d'un grand nombre d'autres pays multilingues, la Russie propose une éducation – primaire pour le moins –, des programmes télévisés et radiophoniques, des ressources internet, des livres et une presse dans quasiment toutes ses langues. La Russie est unique sous un autre aspect également : près de quarante de ces langues indigènes ont un statut de langue officielle. Les langues trouvent toutes un soutien aux niveaux municipal, régional et fédéral.

La Fédération de Russie possède une structure administrative territoriale sophistiquée constituée autour de quatre-vingt-trois entités – cinquante régions, huit territoires, vingt républiques, quatre districts autonomes et une région autonome.

Une région, ou *oblast*, est une entité administrative territoriale caractérisée par une forte domination de l'ethnie Russe et où aucune autre ethnie présente ne compte pour plus d'1 % de la population.

Un territoire, ou *krai*, est une entité administrative territoriale importante qui comprend des districts autonomes où la population composée par les minorités ethniques est dense.

Les républiques sont des entités constituantes dont la population est composée dans des proportions équivalentes par des communautés de Russes et d'autres ethnies. Le nom donné à ces républiques provient de ces dernières. Les républiques ont leur propre constitution et disposent d'une plus grande autonomie vis-à-vis du pouvoir fédéral que les territoires, les régions et les districts autonomes.

Le russe ainsi que la langue du groupe ethnique qui donne son nom à la république, sont toutes deux reconnues comme langues d'État de la république même lorsque ce groupe est une minorité ethnique en son

sein. Dans certaines républiques le statut de langue officielle est également accordé à d'autres langues. Dans ce sens, la préservation de droits équitables pour les langues est un objectif prioritaire et ce par le biais de mesures de protection émanant des constitutions, de lois spécifiques ou d'autres textes prescripteurs à l'échelle des entités fédérales.

Les républiques ont le droit de définir leurs propres langues officielles, et de publier les lois républicaines et fédérales dans ces langues. De même ces autres langues officielles sont sur un pied d'égalité avec le russe lors d'élections, de référendums ou d'activités industrielles, tertiaires ou judiciaires entre autres.

L'établissement de langues officielles n'implique pas l'oubli des langues parlées par les autres peuples habitant la Russie. Chaque entité ethnique dispose d'un droit à préserver sa langue première et à créer les conditions de son enseignement et de son développement. Le Tartarstan est un exemple de république faisant beaucoup d'efforts pour préserver la culture et la langue de ses populations locales Bachkire, Oudmourte ou Tchouvache; De son côté, la Tchouvachie s'efforce à préserver les langues de son territoire, comme le tatar ou le bachkir...

La constitution de la Fédération de Russie garantit à tous ses peuples la préservation de sa langue ethnique, et fait de son apprentissage et de son développement un passage obligé.

Ces principes constitutionnels sont le socle même de la politique gouvernementale russe envers les langues. Ils sont développés et concrétisés au sein de la *Loi sur la langue d'état de la Fédération de Russie*, la *Loi sur les langues des peuples de la République soviétique fédérative socialiste de Russie*, votée durant l'ère soviétique mais encore valide à ce jour, et un certain nombre d'autres lois ayant trait à la culture, à l'éducation etc. La législation en vigueur stipule que toutes les langues ethniques de la Fédération de Russie sont protégées par l'État, qui promeut leur développement, ainsi que le bilinguisme et le multilinguisme, dans tout le pays.

Les citoyens russes n'ayant pas une maîtrise suffisante du russe peuvent participer dans la langue qu'ils parlent couramment aux réunions et conférences au sein des administrations gouvernementales, des services publics, de l'industrie ou encore devant les tribunaux. Un interprète est mis à disposition si nécessaire.

la *Loi sur les langues des peuples de la République soviétique fédérative socialiste de Russie* donne la possibilité d'organiser un enseignement dans toute langue native, quel que soit le nombre de locuteurs du groupe ethnique concerné qui sollicite une telle mise en œuvre. La loi autorise l'usage d'une langue locale aux côtés des langues constitutionnelles dans le cadre de communications officielles effectuées dans des districts ou régions habités par une importante communauté de locuteurs de cette langue.

La *Loi Fédérale sur l'éducation* stipule que les langues d'enseignement et de communication dans un établissement scolaire d'enseignement général doivent être définies par les fondateurs et/ou les statuts de l'établissement. L'État s'engage à accompagner la formation d'experts pour mettre en place un enseignement dans des langues des peuples de Russie qui n'ont ni État, sous la forme d'une république, ni de zone d'autonomie au sein de la Fédération de Russie.

Les autorités régionales Russes sont les plus impliquées dans la mise en œuvre et l'implantation de cette politique étatique ethnolinguistique puisqu'ils rencontrent au quotidien ces problématiques de préservation du multilinguisme ; ces questions les concernent singulièrement.

Prenons comme exemple la situation linguistique et la politique afférente au sein de la république de Sakha (Yakoutie) – et particulièrement la langue yakoute.

SITUATION LINGUISTIQUE, LANGUE ET POLITIQUE ETHNO-CULTURELLE EN RÉPUBLIQUE DE SAKHA (YAKOUTIE)

Composition et expression ethnique

La République de Sakha (Yakoutie) qui couvre une superficie de 3 103 200 kilomètres carrés, est l'entité constituante la plus vaste de la Fédération de Russie. Forte d'une population d'un peu plus d'un million d'habitants, dont 450 000 parlent le yakoute, Sakha (Yakoutie) est une parfaite illustration de ce qu'est un territoire multiethnique en Russie.

Le yakoute est la langue éponyme de l'entité ethnique qui donne son nom à la république (Sakha étant historiquement le nom par lequel se désignaient les Yakoutes). Bien que les Yakoutes soient génétiquement

apparentés aux Mongols, leur langue fait partie des langues turques de la famille des langues altaïques.

Dans la composition ethnique de la population locale, les Yakoutes sont la communauté la plus importante avec 45,6 % d'habitants. L'ethnie Russe vient en seconde position, avec 41,1 % ; les 126 autres groupes ethniques composent les 13,3 % restant. Ces 13,3 % sont composés des minorités aborigènes du Nord – les Evenks, les Évènes, les Youkaguirs et les Tchouktsches. Ils sont regroupés en premier lieu dans 69 colonies à très forte densité de population, principalement dans la région Nord, mais habitent également 238 clans nomades répertoriés.

93,3% de la population de Sakha a une très bonne maîtrise du russe. 87,4 % des Yakoutes, 37,7 % des Tchouktsches, 20,7 % des Évènes, 19,5 % des Youkaguirs et 6,5 % des Evenks considèrent leur langue native comme langue première. Langue minoritaire en Russie, le yakoute est une langue majoritaire dans la République de Sakha (Yakoutie) et possède le statut de langue officielle.

La constitution de la République de Sakha (Yakoutie) garantit le développement inconditionnel du yakoute et du russe au titre de langues d'État, et certaines langues des ethnies indigènes minoritaires de la région du Nord – l'evenki, l'évène, le youkaguir, le dolgane, le tchouktsche, toutes possédant une écriture – au titre de langues officielles.

Un conseil présidentiel sur les politiques linguistiques, dont les fonctions sont de première importance, a été créé au sein de la république. Le gouvernement met en œuvre par ailleurs divers programmes ciblés de développement linguistique. Des événements culturels sont organisés pour promouvoir le dialogue interculturel et améliorer la situation linguistique. Certains jours fériés sont dédiés aux langues indigènes et à la langue russe.

En 2009², mille cinquante neuf livres ont été publiés dans la République de Sakha. 318 l'ont été en yakoute, deux en évène, trois en evenki et cinq en youkaguir. Il y a trente journaux écrits dans la langue yakoute (13 à l'échelle de la république et 17 à celle du district) et douze magazines. Le journal *Tatkachiruk* est publié en evenk. Le journal *Ilken* est écrit en russe, en yakoute et dans d'autres langues minoritaires. Une approche multilingue est utilisée également pour l'almanach *Khalarkhat*. La langue

2 120 000 livres ont été publiés en Russie durant l'année 2009.

russe est utilisée dans 62 % des programmes télévisés diffusés dans la république, le yakoute dans 38 % et les langues minoritaires du nord 1 %. Il y a par ailleurs de plus en plus d'émissions traitant de la culture et de l'histoire des peuples de Yakoutie diffusées en langue russe.

Mécanismes de préservation

Aucune langue ne peut survivre, et encore moins s'enrichir, en dehors de la culture de son ethnie locutrice. Pour cette raison, afin de préserver une langue, il est prioritaire en premier lieu de protéger tous les aspects de cette identité ethnique : culture, coutumes, traditions, folklore, sports traditionnels, cuisine, savoir faire économique et environnement.

Regardons de plus près ce qui est fait pour préserver et développer la culture traditionnelle et la langue yakoute.

La république dispose de cinq cent vingt-huit bibliothèques, soixante-dix-neuf musées, douze théâtres, cinq cent soixante-cinq centres culturels et quatre-vingt-dix écoles d'art pour enfants où sont enseignés musique, peinture, danse et tous types d'arts, dont les arts et artisanats ethniques. La Yakoutie héberge des festivals ethnographiques et organise des expositions d'art itinérantes, ainsi que des tournées d'artistes russes ou étrangers sur son territoire, ou de ses propres artistes dans d'autres régions de la Russie et dans d'autres pays.

Iakoutsk, la capitale de la République de Sakha (Yakoutie), où résident 240 000 personnes, dispose quant à elle des espaces suivants : Le Centre Dramatique National, qui accueille des pièces de théâtre russes ou étrangères jouées en langue yakoute, ainsi que des créations d'auteurs Yakoutes contemporains ; le Centre Dramatique Russe, le Théâtre de Marionnettes et le Théâtre de la Jeunesse, pour n'en citer que quelques-unes ; l'Opéra National, où sont programmés opéras et ballets de compositeurs Yakoutes aux côtés d'œuvres classiques du répertoire mondial ; un orchestre symphonique et l'Ensemble de violonistes de l'École supérieure de Musique de la République de Sakha-Yakoutie ; le Musée National, qui expose des objets reflétant l'histoire et la culture d'origine des Yakoutes et d'autres peuples de la République, ainsi que des peintures et sculptures contemporaines ; le musée ethnographique à ciel ouvert, le Musée du Mammouth,

la Bibliothèque Nationale de la République de Sakha (Yakoutie)³ – la collection la plus importante de livres et documents imprimés en yakoute et dans les langues minoritaires indigènes de la région du Nord⁴ – et dix-huit bibliothèques publiques; l'Académie Nationale de Musique et son internat, où les enfants les plus doués étudient pendant douze ans, peuvent y pratiquer tous les instruments d'un orchestre symphonique en sus des instruments propres aux cultures ethniques de la république; cinq écoles d'art, quatre cinémas; et deux palais des congrès et galeries d'art.

En 2009, la république comptait 654 établissements scolaires, dont 67 % (415) dispensant leur enseignement en langue yakoute pour l'ensemble des disciplines du primaire. Les langues indigènes du Nord sont enseignées en tant que discipline singulière dans 38 écoles (dont certaines nomades)⁵.

Les langues du Nord de la république sont également enseignées à l'Institut des Études et Questions humanitaires des Ethnies Minoritaires du Nord, affiliée à la branche Sibérienne de l'Académie des Sciences de Russie et de l'Institut de Recherche des Écoles Ethniques.

Dans les six établissements d'enseignement supérieur dont dispose la République de Sakha (Yakoutie), les langues minoritaires font également l'objet d'une attention particulière. L'Université fédérale du Nord-Est propose un enseignement en histoire et en philologie dans la langue yakoute (toutes les autres disciplines sont enseignées uniquement en russe). En 2008 et en 2011 le Comité russe du Programme Information pour tous de l'Unesco et le Centre de Coopération Interrégional des Bibliothèques ont organisé, en collaboration avec l'Université et le gouvernement de Yakoutie, deux conférences internationales *Diversité culturelle et linguistique dans le Cyberespace*. En 2010, suite aux recommandations issues de la première conférence, le Centre pour l'étude et la promotion du Multilinguisme dans le Cyberespace a ouvert ses portes sous la responsabilité de l'Université fédérale du Nord-Est.

Le programme *Mémoire de la Yakoutie*, lancé en 2000, prévoit la recherche, la sélection et la numérisation de livres rares en langue yakoute, de documents d'archive en lien avec les aspects primordiaux de l'histoire et de la culture Yakoute ainsi que des enregistrements sonores rares,

3 <http://nlib.sakha.ru>

4 Bibliothèque des minorités ethniques du nord - <http://nlib.sakha.ru/knigakan>

5 Evène (22 écoles), evenki (14 écoles), youkaguir (3 écoles), tchouktche (2 écoles), dolgan (1 école).

historiquement et culturellement importants, d'artistes Yakoutes. Ces œuvres sont mises à disposition sur le site internet de *Mémoire de la Yakoutie*⁶.

En 2005, l'Unesco a fait entrer l'épopée héroïque *Olonkho* dans sa liste des Chefs-d'œuvre du patrimoine oral et immatériel de l'humanité. Ce qui a incité le gouvernement de la république de Sakha à concevoir et mettre en place un programme d'action ciblé, se déployant depuis 2007 jusqu'en 2015, et dont le but est de préserver, développer et faire connaître cette épopée. Entre autres actions, le programme inclut la recherche et le regroupement d'autres épopées, la promotion des conteurs, l'étude de l'*Olonkho* à l'école et la mise en œuvre d'un système d'enseignement spécifique pour apprendre aux enfants l'art de la narration folklorique. L'Université fédérale du Nord-Est demeure le centre névralgique de la recherche consacrée à l'*Olonkho* et du système d'enseignement spécifique à la préservation, l'étude et la popularisation de l'héritage épique dans son ensemble. Le portail <http://www.olonkho.info> donne accès aux textes de l'épopée et plus largement du folklore local en de nombreuses langues.

Des services internet non-textuels se développent également. Des dictionnaires sonores de la langue yakoute sont en cours de réalisation. Plusieurs établissements d'enseignement organisent régulièrement des conférences en ligne sur les différents aspects de la langue et de la littérature Yakoute. D'autres contenus et ressources internet de pertinence équivalente pour la présence de la langue dans le cyberspace sont également élaborés. La langue yakoute bénéficie d'ajouts réguliers dans Wikipédia (<http://sah.wikipedia.org>).

Malgré toutes ces mesures, des études récentes témoignent de la diminution progressive du nombre de locuteurs natifs de la langue yakoute. Dans le même temps, le nombre d'étudiants et de locuteurs du yakoute augmente parmi les membres de l'ethnie Russe.

RECOMMANDATIONS POUR LA PROMOTION DU MULTILINGUISME

Suite à cette analyse d'une expérience singulière au sein de la Fédération de Russie et de l'une de ses entités constituantes, la République de Sakha,

⁶ <http://www.sakhamemory.ru>

nous pouvons tenter de donner une réponse extrêmement générale à la question de ce qu'il faudrait faire pour garantir l'usage d'une langue minoritaire dans un contexte national où existe une langue dominante.

Les contributeurs potentiels à la promotion et au développement d'une langue sont nombreux et variés. Les responsables de ce travail sont les autorités nationales; les autorités locales; le réseau éducatif; les établissements de recherche; les institutions de mémoire; les établissements artistiques travaillant en lien étroit avec des peintres, sculpteurs et architectes locaux; les studios de cinéma; les centres culturels, principalement dans des zones d'habitation isolées, qui remplissent les fonctions d'institution de mémoire et de centres éducatifs et artistiques; les maisons d'édition et les libraires; les médias; l'industrie des technologies de l'information et de la communication; les organisations publiques, les entreprises privées et l'ensemble des individus.

Alors que les TIC pénètrent à grande vitesse dans la vie contemporaine, le développement de la diversité culturelle et linguistique dans le cyberspace, nouvelle sphère d'activités, acquiert une importance toute singulière. Il crée de nouvelles opportunités de préservations des langues et des cultures.

La *Déclaration de Principes* du Sommet Mondial sur la Société de l'Information insiste sur le fait que la politique de construction de la société de l'information devrait être fondée sur le respect de l'identité culturelle, de la diversité culturelle, linguistique et religieuse, et devrait promouvoir le dialogue entre les cultures et les civilisations. Une attention particulière, dans cette construction d'une société inclusive de l'information, devrait être accordée à la création, la circulation et la préservation de contenus dans un grand nombre de langues et sous des formats variés. Le développement de contenu local répondant aux demandes ethniques ou régionales fera la promotion du développement social et économique, et de la participation sociale de tous les acteurs concernés, dont la population des zones rurales, éloignée et marginalisées.

Prenons chacun des contributeurs potentiels séparément afin d'identifier leurs différentes cibles et leur ligne d'action.

Autorités Nationales et Locales

Les politiques et activités des autorités sont de première importance. Une politique efficace inclut une combinaison de mesures destinées à améliorer l'efficacité de tous les autres contributeurs. Les missions (consacrées devant la loi) de ces autres acteurs incluent la promotion de la préservation, de la libre expression et le développement de l'identité non seulement linguistique mais également ethnoculturelle et religieuse des communautés ethniques, la préservation et le développement de leurs valeurs culturelles et de leurs traditions, de tout type de folklore, l'expansion de la sphère d'usage de leur langue native, et la consolidation concrète des principes de pluralisme culturel, de bilinguisme et de multilinguisme. Cet objectif nécessite l'élaboration de lois spécifiques (ou/et l'actualisation d'une législation existante). Le respect de ces lois doit être surveillé et contrôlé.

Les lois fédérales (nationales) et régionales (locales) sur les langues doivent stipuler que l'obtention du statut de langue d'État par certaines d'entre elles ne doit pas empiéter sur les droits d'utiliser leurs langues pour toutes les autres entités ethniques vivant depuis longtemps sur un territoire donné.

Les programmes à mettre en œuvre pour atteindre ces objectifs devraient être fondés sur une logique d'enrichissement intellectuel et culturel mutuel, et protéger les langues minoritaires, les coutumes, traditions et autres valeurs culturelles, ainsi que les institutions reflétant les spécificités culturelles ethniques.

Les autorités devraient contribuer à un enseignement systématique des langues et à la promotion du multilinguisme dans l'éducation, l'administration, le domaine de la loi, de l'éducation culturelle, des médias et du cyberspace.

De nombreuses actions et orientations peuvent permettre d'atteindre ces objectifs :

- la mise en place d'un cadre de régulation pour le développement des langues au niveau national (la Constitution nationale, les Lois fédérales, et les Constitutions des entités constituantes et leur législation) ;

- la conception et la mise en œuvre de stratégies culturelles et éducatives, de politiques et de programmes accentuant explicitement la promotion des cultures et langues minoritaires ;
- des budgets dédiés et d'autres financements en provenance de programmes gouvernementaux pour la préservation et le développement des langues, ou encore des subventions de programmes non gouvernementaux, ainsi qu'une fiscalité modérée et adaptée ;
- l'attribution du statut de langue d'État ou officielle aux langues des minorités ethniques les plus importantes à l'échelle nationale ou au sein des régions à forte population locutrice de ces langues ; autant que possible, l'égalité entre les langues doit être entérinée par la loi.
- la confirmation du statut officiel d'une langue par les services de gestion des documents d'archives des autorités gouvernementales et municipales : l'utilisation de la langue dans les travaux de ces autorités, la publication de textes de lois fédérales ou adoptées au sein d'une république dans cette langue (et la garantie de leur égale force juridique) ; l'utilisation de la langue sur un pied d'égalité avec la langue d'État principale durant les élections, les référendums et les activités industrielles, tertiaires et administratives ;
- la création de bases de données des documents officiels dans cette langue ;
- la création de conseils sur les politiques linguistiques sous l'égide des gouvernements régionaux et/ou nationaux, la détermination de leurs droits et de leurs devoirs ;
- la garantie d'une protection juridique, sociale et économique de la langue par l'exécutif, le législatif et le judiciaire.
- des motivations matérielles pour les experts utilisant langue nationale ou locale dans leur travail ;
- la signature et la ratification des textes internationaux faisant la promotion du multilinguisme ;
- la promotion auprès des entités ethniques concernées de l'intérêt à développer leur langue ;
- le développement des programmes ciblés et des programmes régionaux pour la préservation de la culture et de la langue ;

- un appui et une assistance juridique au développement d'une littérature dans la langue; l'apport d'un soutien, financier entre autres, à la publication de livres aux thématiques variées et de la presse, en particulier des productions écrites adressés aux enfants et à la jeunesse;
- la conception et l'implémentation des stratégies et des programmes faisant la promotion de la lecture dans les langues maternelles et premières;
- la création de partenariats avec des associations culturelles ethniques en dehors des entités administratives dans les territoires où se concentre une forte population de l'ethnie concernée;
- le soutien aux bibliothèques, aux musées, aux archives et autres agences culturelles dans leurs actions de conservation et de développement des cultures et langues minoritaires;
- la mise en place d'écoles spécifiques pour transmettre de génération en génération les expériences des minorités ethniques, leurs traditions, culture et éthique;
- la promotion du réseau éducatif empreint d'une composante ethnoculturelle et son extension dès que cela est possible et nécessaire;
- inclure des cours de littérature et d'enseignement de la langue locale dans les écoles publiques.
- se rapprocher des gouvernements d'autres régions où réside une population importante de locuteurs d'une langue minoritaire, afin de soutenir la mise en place de mesures communes visant à préserver la langue, notamment la fourniture de livres aux bibliothèques publiques et scolaires, la promotion de livres scolaires originaux et des méthodes d'enseignement et d'apprentissage utilisant la littérature, la participation des minorités ethniques dans les programmes de formation de professeurs de deuxième et troisième cycle universitaire;
- l'élaboration d'alphabets et d'écritures pour les langues orales.

Pour ce qui concerne la promotion du multilinguisme dans le cyberspace, les autorités peuvent imaginer à la fois des projets de lois d'ordre général pour créer un environnement favorable et des mesures adaptées à des objectifs spécifiques :

- initier, élaborer et implémenter des programmes de pénétration des TIC;
- dessiner des plans d'action pour promouvoir l'usage public de l'internet, dont des programmes d'alphabétisation informationnelle dans les langues dominantes et minoritaires;
- construire des réseaux de télécommunication dans les zones les plus reculées;
- élaborer des programmes de développement de ressources informationnelles dans les langues minoritaires;
- favoriser la formation aux TIC dans les langues locales;
- promouvoir la création de contenus locaux, leur traduction et leur adaptation;
- promouvoir la traduction de classiques de la littérature mondiale dans les langues minoritaires, celle des auteurs locaux dans d'autres langues, et rendre accessibles ces traductions sur internet.
- établir des réseaux intégrés de ressources informationnelles multilingues;
- présenter des documents numériques et d'archives dans deux langues au minimum;
- promouvoir la recherche et le développement de systèmes d'exploitation, de moteurs de recherche et de navigateurs internet, de dictionnaires en ligne et de livres de référence, ainsi que leur régionalisation, ou localisation⁷.

Centres de recherche (universités et instituts de recherche)

Les centres de recherche fournissent les bases théoriques des actions menées par les gouvernements et autres institutions sociales pour la promotion des langues, et font dans cette sphère à la fois de la recherche fondamentale et appliquée. Leurs responsabilités comprennent :

⁷ Voir dans cet ouvrage : Dwayne Bailey, *Localisation des logiciels : open source et multilinguisme numérique*.

- l'étude des cultures, traditions et pratiques de vie des minorités ethniques ;
- l'étude des langues et leur histoire ;
- l'étude de la situation au présent des langues natives et des difficultés rencontrées ;
- l'étude des politiques et pratiques en faveur de la promotion de langues dans d'autres régions du pays et à l'étranger, et le compte rendu et la dissémination d'expériences pionnières ;
- l'élaboration de propositions pour adapter des expériences de pointe ;
- l'élaboration et la mise en œuvre d'outils pour une évaluation permanente des usages d'une langue par différents groupes sociaux dans les différents domaines où ils l'utilisent ;
- l'élaboration des outils pour mesurer l'efficacité du travail des différentes institutions dont les responsabilités incluent celle de la promotion des langues ;
- soumettre des propositions au gouvernement pour la promotion des langues (des ébauches de textes et programmes de régulation)
- initier et organiser des recherches et conférences pratiques sur les différents aspects que peuvent prendre le développement et la préservation de langues minoritaires ;
- l'établissement de centres de recherche linguistique et culturelle des minorités ;
- la formation d'experts pertinents ;
- donner un visage aux langues et cultures minoritaires auprès du plus large public ;
- l'élaboration des stratégies, des programmes et des plans nationaux de promotion de la lecture – en particulier dans les langues minoritaires – en coopération avec les bibliothèques et les établissements éducatifs, les médias et les éditeurs ;
- l'élaboration de recommandations méthodologiques à l'attention des bureaux et organismes impliqués ;

- concevoir des précis d'orthographe, des méthodes et des dictionnaires bilingues, incluant des enregistrements sonores;
- établir des corpus textuels et des bases de données phonétiques;
- des études consacrées aux champs linguistique et folklorique et appuyées par des expéditions sur le terrain;
- la mise en place d'archives centralisées, incluant des archives électroniques, sur les langues minoritaires;
- acquérir dans la régions ou à l'étranger des archives privées appartenant à des chercheurs et des activistes engagés dans le soutien des langues minoritaires et transmettre ces archives aux institutions de mémoire dépendant de l'État pour leur conservation;
- élaborer des normes et standards d'enregistrement et de représentation de textes, ainsi que des alphabets et écritures pour les langues n'en disposant pas;
- élaborer une langue littéraire unifiée lorsqu'elle n'existe pas;
- documenter les langues minoritaires;
- effectuer des recherches et développer des systèmes d'exploitation, des moteurs de recherche et des programmes de reconnaissance de caractères pour toutes les écritures⁸;
- élaborer des typographies avec des experts d'entreprises spécialisées dans ce domaine.

Éducation

Les établissements d'enseignements primaire, secondaire et supérieur devraient coopérer avec les organismes juridiques aux niveaux fédéral et régional, ainsi qu'avec les institutions de recherche et culturelle, sur les questions du soutien et du développement des langues minoritaires et du multilinguisme. Leurs actions incluent :

- la participation à la conception des règles générales d'enseignement à l'échelle régionale;

⁸ Voir dans cet ouvrage : Pann Yu Mon & Madhukara Phatak, *Moteurs de recherche et langues asiatiques*.

- la formation de professeurs en langues minoritaires pour les écoles et universités;
- la formation d'experts en langues, histoire et culture traditionnelles des minorités ethniques;
- implémenter des programmes de formation d'enseignants pour les troisièmes cycles universitaires;
- concevoir des programmes d'enseignement de base;
- élaborer des programmes pour les universités et des photocopiés;
- élaborer des méthodes d'enseignement des langues et d'amélioration de l'élocution;
- faire des préconisations pour implémenter des nouvelles technologies dédiés à l'enseignement des langues;
- prévoir des programmes universitaires avec des enseignements dans des langues minoritaires;
- utiliser une langue minoritaire comme outil éducatif dans les écoles maternelles (amélioration de l'élocution dans une langue native) ainsi que dans les collèges et lycées;
- accorder des unités de valeur aux étudiants optant pour des cours de langues minoritaires qui ne sont pas leur langue native, ou qui ne l'ont jamais utilisé dans leur enfance, au sein d'établissements d'enseignement situés dans des zones où les minorités ethniques constituent une part importante de la population;
- organiser des cours spécialisés en langues et littératures;
- organiser des compétitions scolaires et universitaires dans les domaines des langues et littératures minoritaires;
- organiser sur une large échelle des conférences et autres événements traitant des problématiques linguistico-culturelles et ethno-culturelles;
- organiser des formations en dehors des campus universitaires (particulièrement aux niveaux interrégional et international);
- organiser des cours de langues minoritaires pour les nouveaux arrivants dans la région ou le pays;

- organiser des camps d'été et des colonies de vacances se déroulant dans une langue minoritaire ;
- organiser des conférences en ligne dans une langue minoritaire ciblée (sur diverses thématiques).

Institutions culturelles

Les institutions et activistes culturels sont d'une importance primordiale pour le soutien des langues – pas uniquement ceux et celles penchés sur la culture écrite et sa préservation, mais les théâtres dramatiques et musicaux, les écoles d'art, les troupes folkloriques, les centres culturels dans des zones périphériques, ainsi que les artistes et individus travaillant dans les milieux culturels.

Bibliothèques, archives et musées

Il est du devoir de ces institutions de préserver, d'entreposer, de populariser et de proposer au public tous les témoignages essentiels de l'histoire particulière d'un peuple.

Les bibliothèques et archives doivent chercher, acquérir, décrire, étudier, faire connaître et entreposer toute publication imprimée, tout enregistrement audio et vidéo réalisés dans une langue donnée, à la fois dans les zones où se concentrent ses locuteurs mais également dans d'autres zones, dont les pays étrangers. En plus des matériels dans des langues minoritaires, toute information à leur propos publiée dans d'autres langues est importante.

Les actions des institutions de mémoire incluent :

- le regroupement, la conservation et l'enrichissement de collections thématiques détaillées de toutes publications et matériels non publiés dans une langue minoritaire ;
- la création de base de données plein texte des journaux et périodiques édités dans la langue concernée ;
- l'établissement une bibliographie exhaustive des sources écrites et imprimées dans la langue en question ;

- l'édition de catalogues bibliographiques de référence des publications dans la langue (particulièrement important pour les langues qui ont développé une écriture récemment) ;
- l'inclusion de fiches descriptives des travaux reflétant l'histoire et la culture d'une minorité ethnique dans les catalogues électronique nationaux de référence de toutes les bibliothèques dans le pays concerné, et dans les catalogues électroniques internationaux ;
- donner de la visibilité à ces travaux – en particulier, en organisant des conférences de lecteurs ou d'autres types de rendez-vous, des clubs de lecture, des rencontres avec des écrivains, critiques, éditeurs, illustrateurs, etc.
- numériser les documents et expositions de musées reflétant l'histoire et la culture d'une entité ethnique, construire des bibliothèques, archives et musées en ligne qui leurs soient dédiés, et permettre un accès public ;
- faire en sorte que les expositions numériques et de musée soient bilingues ou dans la langue ciblée ;
- créer des catalogues numériques dans cette langue au sein des systèmes informatiques des musées ;
- organiser l'archivage des publications électroniques et des expositions sur la diversité culturelle et linguistique, ainsi que des dates et événements d'importance.

En partenariat avec d'autres établissements culturels, éducatifs ou de recherche, les bibliothèques, les musées et les archives peuvent mettre en œuvre différents projets de création de contenu multimédia sur des thèmes traitant des fondateurs des cultures ethniques, du folklore, des collectionneurs, écrivains, artistes, compositeurs et musiciens (dans une large variété de thématique). Des textes, photographies, copies numériques de tableaux et de dessins, des enregistrements sonores et vidéo peuvent être édités sur disques pour une circulation large, et leur version numérique diffusée sur des sites des institutions culturelles, de recherche et éducatives.

Médias

Les médias d'information fédéraux, régionaux et municipaux peuvent devenir des tremplins de la diversité linguistique et culturelle. Des initiatives pourraient être prises pour les aider à :

- préserver et développer des périodiques dans les langues minoritaires, et/ou créer des pages dans ces langues dans d'autres périodiques ;
- organiser des émissions télévisuelles et radiophoniques dans des langues minoritaires (préparer et diffuser de tels programmes, ainsi que des cycles d'émissions sur la vie de la région où ces langues sont utilisées abondamment, et sur la culture d'origine de leurs locuteurs) ;
- mettre en place des sites internet dans des langues minoritaires ;
- ouvrir des portails d'information.

Édition et diffusion de livres

Les maisons d'édition de livres et les libraires peuvent faire la différence dans leur contribution au soutien des langues et du développement du multilinguisme. Lorsqu'une langue n'a pas d'accès à l'industrie du livre, ses locuteurs sont menacés d'exclusion de la vie de la communauté intellectuelle. L'Unesco souligne l'importance de la traduction dans le développement du multilinguisme et le rôle essentiel de l'accès à l'édition pour les langues, promouvant ainsi les échanges de livres et de matériels publiés, ainsi que la libre circulation des idées par voie orale et par l'image.

Les éditeurs peuvent contribuer à la promotion et au développement des langues minoritaires par de multiples initiatives :

- la distribution d'ouvrages de recherche ou de science populaire, ainsi que des fictions (dont des traductions) et des périodiques dans une langue minoritaire.
- la promotion d'œuvres littéraires dans une langue minoritaire, et la recherche de jeunes auteurs parmi ses locuteurs ;
- s'assurer que les bibliothèques des institutions éducatives comportent bien des rayons de livres dans les langues minoritaires ;
- permettre à tous les locuteurs d'une langue minoritaire d'acquérir des livres, particulièrement dans les villages éloignées des centres culturels

et industriels et pour les locuteurs résidant en dehors des zones historiquement habitées par l'entité ethnique en question.

Associations

Les organisations non-gouvernementales peuvent également contribuer à la promotion et au soutien des langues et des cultures. Leurs actions dans ce sens incluent :

- la mise en place d'écoles du week-end, de clubs et d'associations ethnoculturels dans les zones culturelles ethniques autonomes, donnant ainsi accès à des enseignements linguistiques et littéraires.
- l'organisation de compétitions, de festivals et d'autres événements créatifs pour promouvoir les traditions culturelles et linguistiques.
- la participation à l'organisation d'événements spécifiques, comme les Journées de la Langue et de la Culture, à l'intérieur et à l'extérieur de l'ethnie ciblée ;
- la participation à des festivals folkloriques ;
- l'établissement de contacts avec des compatriotes résidents en dehors de la région ou du pays.

Initiatives privées

Les langues doivent trouver du soutien non seulement de l'État, de ses institutions ou d'organisations non-gouvernementales, mais également des individus et des groupes d'individus qui peuvent participer à leur défense et promotion :

- initier et enrichir des espaces linguistiques dans les langues minoritaires au sein de Wikipédia ;
- créer et actualiser des sites web, des blogs, des comptes Twitter et des réseaux sociaux dans les langues minoritaires⁹.

⁹ Voir dans cet ouvrage : Vassili Rivron, *L'usage de Facebook chez les Éton du Cameroun*

L'industrie des TIC

L'industrie des TIC doit garantir sa participation dans le soutien et l'amélioration du statut des langues. Elle peut concentrer ses actions dans les domaines suivants :

- l'élaboration de normes techniques qui tiennent compte des demandes des minorités ethniques et la promotion de leur mise en œuvre¹⁰ ;
- l'élaboration de typographies numériques pour les langues ethniques minoritaires ;
- une participation à l'élaboration des normes internationales Unicode et l'implémentation d'un agencement de clavier unifié ;
- la localisation des logiciels diffusés et la création de nouveaux logiciels libres pour soutenir les langues locales ;
- l'élaboration de langages de modélisation informatique et de systèmes de traduction automatique¹¹ ;
- la promotion de l'usage de langues minoritaires dans les courriers électroniques, les messageries instantanées et les utilitaires de gestion de messages ;
- l'élaboration de méthodes d'apprentissage de l'informatique et de dictionnaires ;
- l'établissement de noms de domaines et d'adresses de courriers électroniques multilingues¹² ;
- l'élaboration et l'implémentation de logiciels prenant en compte des noms de domaines et des contenus internet multilingues ;
- l'établissement de moteurs de recherches localisés et adaptés à chaque langue minoritaire ;
- la mise en œuvre de sites et de portails d'informations bilingues ;

10 Voir dans cet ouvrage : Stéphane Bortzmeyer, *Le multilinguisme dans la normalisation de l'internet*.

11 Voir dans cet ouvrage : Joseph Mariani, *Les technologies de la langue en soutien au multilinguisme*

12 Voir dans cet ouvrage : Stéphane Bortzmeyer, *La gouvernance de l'internet à l'heure du multilinguisme*

- le transfert des ressources informationnelles vers les médias numériques;
- le développement des ressources non-textuelles sur l'internet (comme les communications vocales VoIP, les sites de streaming et de vidéo à la demande)¹³.

L'ensemble de ces mesures peut apporter les effets désirés dans la mesure où l'ethnie dans son ensemble – et pas seulement son élite intellectuelle, culturelle et dominante – produit des efforts intellectuels et émotionnels importants, et affiche sa bonne volonté, son désir et son intérêt pour que survive et se développe sa culture unique fondée sur sa langue.

BIBLIOGRAPHIE

[KUZMIN, EVGENY (ED). 2008] “Report by the Russian Federation to Unesco General Conference on Measures Taken to Implement the Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace”, In: *Preservation of Linguistic Diversity: Russian Experience*. Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre), pp. 8–36.

[KUZMIN, EVGENY AND PARSHAKOVA, ANASTASIA. (EDS). 2011] *Развитие многоязычия в киберпространстве: пособие для библиотек* (Promoting Linguistic Diversity in Cyberspace: A Handbook for Libraries). Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre).

[KUZMIN, EVGENY AND PLYS, EKATERINA (EDS). 2007] *Языковое разнообразие в киберпространстве: российский и зарубежный опыт* (Linguistic Diversity in Cyberspace: Russian and International Experience). Collection of analytical works. Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre).

[— 2008] “Development of Multilingualism on the Internet as a New Field of Activity of the Russian Committee of the Unesco Information for All Programme and the Interregional Library Cooperation Centre”, In: *Preservation of Linguistic Diversity: Russian Experience*. Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre), pp. 65-85.

[— 2008] *Представление языков России и стран СНГ в российском сегменте Интернета* (Presentation of Languages of Russia and other CIS Countries in the Russian Segment of the Internet). Collection of reports. Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre).

13 Voir dans cet ouvrage : Tunde Adegbola, *Multimédia, langue des signes, langues écrites et orales*.

[——— 2008] *Preservation of Linguistic Diversity: Russian Experience*. Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre).

[——— 2010] *Языковое и культурное разнообразие в киберпространстве* (Linguistic and Cultural Diversity in Cyberspace). Proceedings of the international conference held in Yakutsk, Russian Federation, 2-4 July, 2008. Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre).

[——— 2011] *Linguistic and Cultural Diversity in Cyberspace*. Proceedings of the International Conference held in Yakutsk, Russian Federation, 2-4 July, 2008. Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre).

[KUZMIN, EVGENY, PLYS EKATERINA, KISLOVSKAIA, GALINA AND TCHADNOVA, IRINA (EDS). 2008] *Многоязычие в России: региональные аспекты* (Multilingualism in Russia: Regional Aspects). Moscow: Межрегиональный центр библиотечного сотрудничества (Interregional Library Cooperation Centre).

MULTIMÉDIA, LANGUES DES SIGNES, LANGUES ÉCRITES ET ORALES

Les systèmes d'écriture n'ont pas été développés pour toutes les langues en même temps, ce qui confère aux langues écrites un grand avantage. Alors que nous avançons plus profondément dans l'âge de l'information, comment faire en sorte que les inégalités des âges agraire et industriel ne soient amplifiées à notre époque informationnelle ? Si l'alphabétisation est une priorité en tant que valeur fondamentale du monde moderne, comment utiliser Internet pour permettre l'expression, la mémoire et l'extension de leur champ de connaissances aux locuteurs des langues non-écrites ?

Article original en anglais.
Traduit par Laurent Vannini.



TUNDE ADEGBOLA, chercheur, consultant et activiste culturel, possède une longue expérience dans les champs de l'information et des médias de communication. Directeur général de l'African Languages Technology Initiative, où une équipe de chercheurs adapte les technologies linguistiques aux langues africaines.

TUNDE ADEGBOLA

MULTIMÉDIA,
LANGUES DES
SIGNES, LANGUES
ÉCRITES ET
ORALES

L'invention de l'écriture fut une étape essentielle du développement humain, facilitant la documentation détaillée et précise des expériences et idées au-delà des confins du cerveau. Ce développement favorisa la mobilité de ces expériences et idées, les libérant de leurs contraintes spatiales et temporelles, et permit dès lors leur partage entre différentes cultures dans le monde ou au sein de chacune d'entre elles. Cependant, toutes les langues n'ayant pas créé des systèmes d'écriture au même moment, l'écriture donne un avantage aux langues qui en disposent; celles qui en sont dépourvues restant relativement limitées dans leur capacité et leur rapidité pour partager les expériences et idées de leurs locuteurs. Les récents développements dans le domaine des technologies numériques ont simplifié la création de ressources informationnelles et cognitives pour des langues sans système d'écriture. Au cœur des technologies modernes de l'information et de la communication, le multimédia est devenu l'un des symboles les plus remarquables de l'ère de l'information en ce sens qu'il facilite et enrichit la communication à la fois dans les langues orales et écrites mais également dans les langues des signes¹. Le cyberspace revêt dès lors la dimension prometteuse d'un territoire de communication véritablement inclusif.

Plus nous progressons dans cette ère de l'information, plus il devient nécessaire de nous assurer que ne s'y reproduisent pas les formes d'inégalités qui ont caractérisé les âges agraires et industriels. Tandis que d'un côté nous poursuivons notre quête d'une alphabétisation généralisée comme valeur fondamentale du monde moderne, de l'autre nous devons faire en sorte que les espaces auxquels donnent accès les technologies numériques ne soient pas utilisables uniquement pour les locuteurs de langues écrites.

1 Voir dans cet ouvrage: Annelies Brafort & Patrice Dalle, *Accessibilité du cyberspace: Langues des signes*.

Il est impératif de développer des techniques et outils multimédias qui s'emparent des avantages des technologies numériques modernes afin de rendre accessibles et utiles dans le cyberspace les langues sans système d'écriture et de rendre possible l'utilisation des langues des signes.

LANGUES ORALES, ÉCRITES ET SIGNÉES.

Le langage est un système sémiotique dans lequel des règles associent des symboles à du sens. En tant que système de communication, le langage a pour caractéristique l'agencement d'un ensemble fini de symboles visuels ou auditifs selon un ensemble fini de règles, tout en donnant lieu à la possible production d'un ensemble infini d'énoncés. Cette capacité de production d'un ensemble infini d'énoncés à partir d'un ensemble fini de symboles et de règles fournit au langage le socle permettant de faire référence à un vaste ensemble d'objets simples, de décrire des concepts labyrinthiques ou d'exprimer des idées complexes.

Dans le discours, le langage est rendu possible par l'utilisation de symboles auditifs. Le son est une conséquence de la compression et de la raréfaction de l'air dans le temps, aussi y a-t-il une dimension temporelle stricte à l'usage traditionnel de signaux auditifs pour donner vie au langage. Lorsque nous utilisons la parole pour mettre en œuvre le langage, les phénomènes auditifs produits déclinent et deviennent imperceptibles en très peu de temps. Pourtant, le plus souvent, les informations transmises par le biais de ces activités auditives restent valables pour une longue durée, qui va bien au-delà de la durée de vie extrêmement courte des phénomènes auditifs produits pour les représenter. Cette restriction temporelle eut des conséquences sur la capacité à documenter et réutiliser l'information et les connaissances exprimées dans le discours.

Dans l'écriture, le langage se matérialise au travers de l'utilisation de marquages visuels sur des supports appropriés. Ces marquages peuvent être des codes qui représentent les concepts à communiquer directement, ou par la représentation indirecte des sons exprimés dans les énoncés discursifs décrivant eux ces concepts de manière directe. La durée de vie de tels marquages visuels est généralement sans commune mesure avec celle restreinte des sons produits dans le discours. Cet avantage temporel du mot écrit sur le mot parlé est un facteur important du développement de l'écriture.

Dans les langues des signes par contre, les symboles visuels sont produits au présent pour compenser les déficiences auditives des personnes sourdes ou malentendantes. Les symboles visuels des langues des signes étant produits dans une temporalité, ils sont également sujets à des restrictions temporelles similaires à celles caractérisant le discours.

Au-delà des dimensions temporelles qui distinguent langues orales, écrites et signées, il existe plusieurs autres différences d'importance entre ces moyens de communiquer. D'un point de vue spatial, la portabilité du média sur lequel est enregistrée l'information permet aux informations et idées écrites de voyager loin des sources dont elles proviennent. Au contraire, des langues exclusivement orales s'appuient principalement sur des processus de mémorisation et de récitation, ce qui restreint leur utilisation à un mode de présentation en présence directe.

Un grand nombre de langues de par le monde ont développé des systèmes d'écriture mais un nombre plus grand encore restent essentiellement dans la sphère de la parole. Les cultures qui utilisent des langues écrites ont bénéficié pleinement des propriétés spatiales et temporelles de l'écriture. Le partage des informations et connaissances via l'écriture et la lecture de textes permet à ces contenus de se déplacer aussi loin que le support écrit l'autorise. Au sein des cultures qui utilisent toujours des langues non écrites, chaque personne doit apprendre principalement par la mémorisation et la récitation. Dès lors le périmètre de diffusion d'une connaissance codée dans une langue orale est déterminé par la puissance de propagation du bouche à oreille. De plus, en raison des limites de la mémoire humaine en termes de pertinence des éléments remémorés, l'information et la connaissance codées dans les langues non écrites sont vouées à dépendre des contraintes mnésiques humaines.

LANGUES, INCLUSION ET LE DÉFI DE L'ALPHABÉTISATION.

Les avantages de l'alphabétisation ont donné une longueur d'avance aux cultures qui utilisent des langues écrites, et celles qui s'appuient sur des langues sans écriture en sont réduites à rester à la traîne. De surcroît, des études ont montré une corrélation causale entre alphabétisation et développement humain, et l'alphabétisation est devenue l'un des indices du développement des sociétés humaines, ce qui conduit à des efforts

conséquents dans le but d'augmenter le nombre de personnes alphabétisées partout dans le monde.

Du fait de cette corrélation causale, les personnes analphabètes appartenant à des cultures de langues écrites bénéficient d'une aide importante à l'alphabétisation, tandis que dans le même temps, dans les cultures de langues non-écrites, les efforts se concentrent autour du développement de systèmes d'écriture. Néanmoins, en dépit des énergies et moyens mis en œuvre pour améliorer les niveaux d'alphabétisation, le nombre de personnes analphabètes est toujours élevé sur la planète, et ces dernières continuent à vivre avec une capacité réduite à participer pleinement au processus de développement sociétal et humain.

Si des individus ne sont en mesure de communiquer leurs idées qu'aux personnes appartenant au même environnement géographique, ils perdent la capacité à diffuser ces idées dans des espaces éloignées tant culturellement que physiquement. Une incapacité qui a deux conséquences majeures. D'un côté, cette limitation est subie par des populations qui pourraient bénéficier de la propagation de ces idées, pendant que d'un autre côté, de telles idées ne sont pas enrichies des contributions et affinements d'un nombre suffisant de personnes qui auraient été en mesure de les améliorer et de les perfectionner pour le bénéfice de l'humanité tout entière. La compréhension humaine de l'environnement naturel conduit à une aptitude à prédire et contrôler le comportement des éléments. Une aptitude qui a d'importantes implications pour le développement humain. Le degré de compréhension de l'environnement naturel par les hommes dépend fondamentalement de la capacité à construire des idées à partir d'autres idées, et des connaissances à partir d'autres connaissances. Par conséquent, la capacité à échanger et partager des informations et des connaissances est primordiale pour le développement de l'homme.

Alors que la communauté mondiale est de plus en plus connectée et que le système économique global tend à devenir un marché commun unique, les personnes dont les langues n'ont pas d'écriture continueront de souffrir des conséquences des limites temporelles et spatiales de leur idiome, laissant aux populations de locuteurs de langues écrites la part belle. Une communauté mondialisée où se développeront de plus en plus ces inégalités.

MULTIMÉDIA ET MULTIMODALITÉ

Les développements issus des technologies de l'information et de la communication (TIC) ont transformé notre monde et continuent à le modifier dans des proportions qui n'ont pas de précédent. Ces techniques ont eu un impact phénoménal sur la manière dont nous communiquons, apprenons et appréhendons les connaissances. Le multimédia est l'une des formes nouvelles de présentation de l'information issue des évolutions des TIC. Le terme multimédia désigne une combinaison de différentes formes d'expression du contenu, incluant le texte, le son, l'image fixe, la vidéo, l'animation ou encore le contenu interactif. Il se distingue des médias traditionnels comme les livres dans la mesure où il peut solliciter plusieurs modes de perception, auditif, visuel ou encore tactile, et ce simultanément ou séquentiellement.

Cette capacité à présenter l'information par le biais de médias différents faisant appel à plusieurs sens humains est probablement la caractéristique essentielle du multimédia au regard de notre communication, de notre enseignement et de notre approche de l'information. Une propriété qui favorise non seulement l'accès à l'information, mais contribue également à améliorer notre compréhension.

Les livres se composent traditionnellement uniquement de textes et d'images fixes utilisées dans la plupart des cas pour illustrer l'écrit. Le livre comme support pour la gestion de l'information et des connaissances est en principe limité à ces formes de médias. Dans le cas du multimédia, le contenu peut être présenté sous une forme écrite articulée avec des images fixes mais également avec du son, de la vidéo et de l'animation. Le son peut prendre la forme d'un discours, d'une musique ou de bruitages tandis que la vidéo peut présenter des séquences d'images animées qui ne sauraient être visionnées par le lecteur/spectateur en d'autres circonstances. L'animation rend possible l'illustration visuelle de phénomènes qui ne pourraient normalement être qu'imaginés, ouvrant ainsi la voie à une perception plus profonde de phénomènes jamais vécus empiriquement.

Le texte écrit, les images fixes, la vidéo et l'animation sollicitent la vue, tandis que la parole, la musique ou les bruitages impliquent l'ouïe, et cette combinaison des sens induit une meilleure mémorisation pour les utilisateurs de ces informations. De plus, le média interactif engage ces derniers dans des échanges et leur donne la capacité de contribuer à

l'information, faisant d'eux des acteurs de l'information et pas uniquement de simples spectateurs consommateurs d'information. Du fait de la sollicitation multimodale des sens humains induite par le multimédia, une information qui pourrait paraître cryptique dans un médium, semblera plus facilement déchiffrable codée dans un autre médium. De même, l'information difficile d'accès dans un médium sera plus aisée à transcrire dans un autre médium. Ce qui a véritablement révolutionné notre façon de consommer de l'information et de partager nos connaissances.

DES RESSOURCES DOCUMENTAIRES NON ÉCRITES

Aussi importante que soit la révolution symbolisée par cette implication multimodale et multimédia des sens, on peut regarder de façon encore plus significative la dimension du multimédia comme outil de documentation de l'information. À ce titre, c'est un outil qui ne fait pas de distinction préférentielle entre présentation visuelle ou description sonore de l'information, et qui peut être développé dans un grand nombre de médias. Et une telle surface de possibles conjure tout préjugé contre telle langue ou tel dialecte. L'écriture en tant que médium d'information est un système d'encodage de l'information, tandis que d'autres médias dans le matériau composite qu'est le multimédia produisent des représentations directes et intuitives des objets et concepts qu'on leur demande de représenter. Les images ne demandent pas à être déchiffrées et un discours enregistré est accessible immédiatement aux auditeurs sans l'obligation préalable d'apprendre à décoder l'information comme c'est le cas avec le texte écrit.

Comme outil de documentation de l'information, un système d'écriture a besoin d'être développé spécifiquement pour une langue donnée. Même s'il existe des cas de langues ayant adopté et adapté des systèmes d'écritures provenant d'autres langues, les adaptations nécessaires engendrent des distinctions telles que le système d'écriture ainsi créé devient unique sous bien des aspects. Aussi, le système d'écriture d'une langue n'est pas utilisable directement par une autre langue, et le développement d'un tel système, quelle que soit la langue, exige la mise en place de normes. De telles normes requièrent des efforts structurés et volontaires. Une culture qui ne serait pas capable de s'organiser pour développer de telles normes

aura sans doute des difficultés à développer un système d'écriture accessible à tous et utile pour tous. Certains éléments du matériau composite multimédia ne sont pas sujets à ce type de problèmes.

Les efforts délibérés consentis pour apprendre à lire et à écrire ne doivent pas être négligés. Lorsqu'un tel processus est initié suffisamment tôt au cours d'une vie, l'alphabétisation rentre dans le cadre d'un processus plus large d'intégration. En revanche, dans le cas d'adultes analphabètes, apprendre à lire et à écrire a des allures de montagne à gravir. Même si les programmes d'alphabétisation abondent dans le monde, et malgré les efforts incessants d'organisations nationales et internationales tels que l'Unesco, il existe encore et toujours des écarts d'alphabétisation patents de par le monde. Dès lors, même si l'alphabétisation est devenue une condition essentielle de l'expérience humaine, la capacité unique de documenter cette expérience humaine grâce au multimédia devrait être exploitée comme l'un des bénéfices de l'ère de l'information et ce particulièrement pour les analphabètes, mais également pour les utilisateurs langues non écrites ou signées.

Pour autant l'utilisation du multimédia n'est pas entièrement chose nouvelle pour l'humanité. La plus grande partie de la préhistoire s'inscrit dans une quête poursuivie par nos ancêtres de documenter leurs expériences afin de les offrir aux générations suivantes. De telles aspirations à laisser des traces pour les générations futures sont parfois exprimées consciemment, mais elles ont été la plupart temps des aveux non conscients, présents dans les empreintes d'un autre temps, qui nous ont renseignés sur leurs façons de vivre et leurs expériences. C'était là leur seule option avant l'invention de l'écriture. Peintures rupestres, sculptures antiques et de nombreux autres artefacts préhistoriques peuvent sérieusement être imaginés comme des tentatives primitives de documenter de l'information en multimédia. Cependant, en raison des limitations propres aux technologies utilisées alors, ces tentatives se cantonnaient dans un seul médium visuel. Les technologies numériques contemporaines nous ont quant à elles donné la capacité d'exploiter multimédia et multimodalité d'une manière qui aurait été inimaginable à l'époque.

Aussi, pour soutenir nos exigences d'amélioration du niveau d'alphabétisation dans les communautés humaines où le besoin est le plus sensible et pour développer des systèmes d'écriture pour des langues orales, nous devrions intégrer les rôles que pourraient y jouer le multimédia et les

nouvelles voies ouvertes par les TIC modernes dans la documentation de l'information et la diffusion des connaissances.

DE L'ANALPHABÉTISME À L'ALPHABÉTISATION.

Les arguments présentés ci-dessus n'ont pas pour objet d'insinuer que le multimédia n'est pas d'ores et déjà utilisé à des fins productives. Ils sont au contraire avancés pour inciter à une utilisation plus créative encore du multimédia dans le cyberspace, afin d'élargir le champ de son usage à des analphabètes et à des utilisateurs de langues non écrites ou signées. L'actuelle prépondérance de l'information écrite dans le cyberspace est liée non seulement à l'efficacité et à la portabilité des textes écrits, mais également à la popularité que l'écriture avait acquise comme médium d'information essentiel avant l'avènement de l'internet. Pour cette raison, le cyberspace s'est développé dans une écologie du mot écrit, et reflète avant tout ce biais générique pour l'écriture. Pour autant, il existe des alternatives multimédias qui peuvent être utilisées pour compléter l'écriture dans le cyberspace.

Nous pouvons aborder sous deux angles différents le multimédia en tant que possible porte d'entrée dans le cyberspace pour des personnes analphabètes et les locuteurs de langues non-écrites ou signées. En premier lieu, nous pouvons l'utiliser dans ses fonctions basiques, par exemple pour enregistrer et faire réécouter de l'information sous forme de discours, ou par le concours de dessins et l'affichage d'images. En second lieu, nous pouvons exploiter des technologies plus complexes, à l'instar des technologies de la langue, et en particulier celles dédiées au traitement automatique du langage naturel pour accompagner l'alphabétisation.

De nombreuses formes d'usage basique du multimédia peuvent bénéficier aux deux populations précitées. Avec d'un côté la popularité des techniques d'enregistrement sonores utilisant des algorithmes de compression réduisant la quantité de données, comme le MP3, et de l'autre la réduction constante du coût des médias d'enregistrement pour un résultat toujours plus fiables, il est désormais possible d'enregistrer de nombreuses heures de discours par le biais d'équipements mobiles, accessibles et de bonne qualité. De tels enregistrements sont des documents pertinents pouvant être utilisés à différentes fins comme le sont généralement les textes écrits. Ils peuvent être réécoutés chaque fois qu'il est nécessaire de consulter

les contenus et ils peuvent être indexés, facilitant ainsi une récupération efficace d'information. De la même manière, les images fixes, la vidéo et l'animation peuvent être utilisées pour raconter des histoires, soit comme compléments à des séquences enregistrées de discours, soit séparément et se suffisant à eux-mêmes.

À un autre niveau d'usage, les technologies de la langue peuvent également accompagner une démarche d'alphabétisation en ligne. Malgré l'efficacité et la portabilité de l'écriture, la parole reste encore le moyen de communication privilégié des humains. L'organisation coûteuse de conférences internationales mettant en vis-à-vis interactif des intervenants proposant leurs réflexions et une audience, le développement de livres audio ou d'autres technologies liées à la parole témoignent toutes de l'importance de la parole pour l'humanité. Par conséquent, les investissements visant au développement des technologies vocales se justifient-ils par cette préférence donnée par l'homme à la parole. De tels investissements irrigueront inévitablement les utilisations dédiées à l'accompagnement de l'alphabétisation. Toutefois ce ruissellement devra être délibérément collecté et exploité au bénéfice des analphabètes et des locuteurs de langues non-écrites ou signées. Les techniques de Reconnaissance Automatique de la Parole par exemple peuvent être utilisées pour convertir des idées exprimées par une personne analphabète en texte écrit, et celles de synthèse vocale à partir du texte peuvent lui permettre de découvrir des textes écrits. Ainsi cette personne aura été en mesure d'interagir avec la littérature sans pour autant savoir lire ou écrire.

L'alphabétisation est définie comme la capacité à lire et à écrire, et son importance est liée à ce qu'elle permet un accès sans précédent à l'information et à la connaissance. Au regard des niveaux actuels d'accessibilité au multimédia et des développements dans les technologies de la langue, dont celles de reconnaissance vocale, couplés à des coûts de plus en plus bas, il est nécessaire de reconsidérer la valeur relative accordée au texte écrit et de donner plus d'espace pour les analphabètes et les utilisateurs de langues non-écrites ou signées.

L'ensemble de ces techniques, qu'elles soient éléments basiques du multimédia ou composantes plus complexes, se rencontrent fréquemment dans le cyberspace. Jusqu'à présent cependant, elles n'avaient pas été considérées comme des moyens viables de communication pour les analphabètes en raison des coûts élevés associés à l'utilisation de ce type de

média dans l'ère pré-internet. À l'ère du numérique, de tels usages sont désormais de plus en plus abordables. Bien qu'ils puissent encore paraître relativement onéreux dans le cadre de programmes d'alphabétisation de masse, ils sont voués à devenir de plus en plus accessibles à mesure que nous progressons dans l'âge de l'information.

Le taux actuel de pénétration des téléphones mobiles dans des zones urbaines et rurales des pays en voie de développement est un indicateur patent de la possibilité de diffusion du multimédia en vue d'atteindre à la fois les populations analphabètes et celles locutrices de langues non-écrites ou signées. Il est donc essentiel d'appréhender le multimédia sous l'angle de telles perspectives futures afin d'exploiter pleinement ses avantages dans la mise en œuvre d'un cyberspace inclusif.

LE MULTIMÉDIA POUR UN CYBERSPACE INCLUSIF

Avant la révolution informationnelle, le monde avait déjà traversé une révolution agraire et une révolution industrielle. Chacune de ces révolutions se caractérisait par des changements abrupts et intenses. Inévitablement dans de telles situations, les êtres humains ne peuvent être que submergés et beaucoup ne peuvent pas suivre, souffrant plus encore d'inégalités.

Afin de réduire les niveaux d'inégalité qui émergent dans le sillage de la révolution informationnelle, le cyberspace, en tant que l'un des produits les plus importants de ce bouleversement, se doit d'être le plus inclusif possible. Il doit être le plus ouvert afin de permettre à toute personne quelle que soit son origine, son degré d'alphabétisation ou sa langue, écrite ou non, d'y contribuer et d'en bénéficier.

La fracture numérique est devenue l'une des métaphores importantes pour décrire les niveaux d'inégalité dans l'âge de l'information. Selon la définition généralement acceptée, la fracture numérique est *« le fossé séparant des personnes ayant un accès réel aux technologies numériques et de l'information et celles ayant un accès restreint ou aucun accès. Elle rend compte à la fois du déséquilibre en termes d'équipements matériels et d'accès physique, et des ressources et compétences nécessaires pour participer réellement en tant que citoyen numérique »*². Même si la description abonde

² http://www.africa4all-project.eu/index.php?option=com_glossary

en exemples de freins à la possible contribution en tant que citoyen du cyberspace, la connotation généralement liée à la fracture numérique lorsqu'elle est évoquée relève plus de l'absence d'accès physique aux technologies de l'information et numériques. Ce qui est dès lors suggéré est de l'ordre d'une carence en équipements et en bande passante.

Mais aussi importante soit cette variable, l'idée que l'absence d'accès physique est la principale problématique au cœur de la fracture numérique est trompeuse. Nous savons aujourd'hui que si toutes les contraintes technologiques à la participation de tous en tant que citoyens numériques étaient levées, subsisteraient encore des freins linguistiques et des difficultés culturelles. Certains de ces facteurs peuvent être minimisés à l'aide du multimédia et des technologies de la langue, parce qu'ils fournissent un champ beaucoup plus large d'action pour communiquer l'information.

En ayant en tête que la révolution informationnelle est propulsée par les technologies numériques, elle dispose de ce fait d'une capacité propre à relever le défi de l'inclusion et d'œuvrer à la réduction d'une grande partie des inégalités ayant accompagné son avènement. Il nous incombe à nous autres humains d'identifier les besoins et de concevoir les programmes qui rendront cette dynamique possible.

CYBERACTIVISME ET LANGUES RÉGIONALES DANS LES PRINTEMPS ARABES DE 2011

Le Web n'est pas écrit ; il s'écrit chaque jour. Des millions d'individus s'emparent d'outils aussi différents et complémentaires que des sites de mise en réseau ou des plateformes d'échange de fichiers. Des pages personnelles au Web dit 2.0, quel rôle les technologies jouent-elles ? Comment favorisent-elles l'expression individuelle et l'intelligence collective ? Quel rôle jouent les langues dans ce redéploiement numérique des liens entre les êtres humains ? Entre globalisation et protection des langues locales, faut-il sacrifier sa langue pour être entendu de la communauté globale ?

Article original en anglais.
Traduit par Laurent Vannini.



ADEL EL ZAIM est directeur général du Service des Relations Internationales à l'Université de Sherbrooke, Québec, Canada. Il est linguiste et expert en TIC et leur rôle dans le développement. Auparavant, il a travaillé au Centre de recherche pour le développement international Canada (CRDI) et pour divers centres de liaison et de transfert au Québec.

ADEL EL ZAIM

ARABES DE 2011
LES PRINTemps
RÉGIONALES DANS
ET LANGUES
CYBERACTIVISME

2010 et 2011 resteront dans l'histoire comme deux années de rupture brutale au Moyen Orient et en Afrique du Nord. Néanmoins l'origine du « Printemps Arabe » est plus ancienne. En réalité, 2011 marque le point culminant et le succès, en Tunisie et en Égypte, de révolutions qui avaient été amorcées en plusieurs occasions durant les premières années du XXI^e siècle. Les révoltes ont ensuite essaimé dans d'autres pays, certaines se soldant par le massacre de la population (Syrie), d'autres par une guerre civile et l'intervention internationale (Libye), d'autres encore par une répression silencieuse (Bahreïn).

Si les jeunes du Moyen Orient et d'Afrique du Nord sont souvent désignées comme étant les architectes de ces victoires insurrectionnelles, le rôle joué par la jeunesse a un impact similaire partout sur la planète. Que ce soit en Chine, en Birmanie ou en Iran, au Maroc, en Tunisie ou en Égypte, en passant par la Libye, le Yémen, Bahreïn ou la Syrie en 2010 et 2011, sans oublier l'Espagne, la Grèce, le Sénégal, ou plus récemment encore les États-Unis ou la Russie, une constante se dessine : celle d'une jeunesse debout sur la ligne de front. Les révoltes, mais surtout les répressions qui les accompagnent, ne cessent de coûter des vies et de ruiner des économies. Dans un grand nombre de pays, comme au Sénégal¹, la population se soulève contre la corruption et les stratégies présidentielles pour rester à jamais au pouvoir.

La question du rôle des Technologies de l'Information et de la Communication (TIC) et des médias sociaux dans les mouvements insurrectionnels en Égypte et en Tunisie a été posée très tôt. Certains chercheurs et journalistes se demandèrent si les soulèvements devaient être désignés sous le terme de « Révolution Numérique » ou encore de « Révolution 2.0 ». D'autres s'interrogèrent plus sérieusement sur les

1 <http://www.senrevolution.com>

raisons de leur localisation temporelle et géographique. Pourquoi maintenant? Pourquoi en Tunisie, en Égypte, et quel pays serait le suivant? Ou encore comment mettre à profit les TIC et s'appuyer sur la capacité de la jeunesse à les utiliser, pour réformer un pays, pour créer de nouveaux systèmes politiques, de nouvelles catégories d'hommes et de pratiques politiques?

Des dizaines de livres et d'études scientifiques ont été publiées en Égypte, en Tunisie, et ailleurs, à propos des soulèvements ayant ébranlé ces deux pays. Des ouvrages se livrant à l'analyse de cette nouvelle ère de renaissance et de ses réformes, de ses conséquences, de ses retombées positives, de son histoire et de ses modèles. Des produits culturels dérivés, incluant des chansons, de la musique, des films et des photographies, continuent d'être produits et de proliférer de par le monde.

Mais qu'en est-il des langues? La relation entre langue et révolution est un vaste champ d'étude qui peut être abordé sous des perspectives très différentes. Témoin direct des mouvements d'insurrection en Égypte au printemps 2011, j'ai été particulièrement intéressé par le rôle de la langue dans les événements, et plus spécifiquement son usage dans les médias sociaux égyptiens. J'ai étudié comment elle était utilisée par les protestataires et les mouvements politiques pour communiquer, mobiliser et organiser des actions; ou encore pour documenter et partager des moments clés de la répression, de la victoire et de la célébration. À partir de ces observations, j'ai pu étudier l'impact de la révolution sur la présence de la langue arabe sur le web et dans les médias sociaux. Je formule l'hypothèse que les révolutions dans les pays arabes, dès leurs premiers balbutiements, ont accéléré la production et l'utilisation de contenus en langue arabe, et ce particulièrement dans les médias sociaux. Ces contenus sont l'œuvre d'individus, d'organisations et de médias et améliorent le rang de l'arabe dans le classement des langues les plus utilisées en ligne. Un certain nombre d'autres recherches viennent aujourd'hui corroborer mes conclusions. Pour autant, le travail d'analyse ne fait réellement que commencer.

L'USAGE DES MÉDIAS SOCIAUX DANS LES PAYS ARABES

Les médias sociaux, tels que Facebook, Twitter ou YouTube, font partie des outils internet les plus utilisés dans les pays arabes. Facebook est un outil de communication de masse d'une efficacité telle qu'il est perçu comme une menace, et qu'il a de ce fait été bloqué, puis débloqué, à de nombreuses reprises par les gouvernements de différents pays allant de la Tunisie à la Syrie au cours des trois ou quatre dernières années. Ces manœuvres avaient pour but de contenir la contestation sociale et politique.

La première édition² du *Rapport sur les médias sociaux dans le monde arabe*³ dénombrait, au 5 janvier 2011, plus de vingt et un millions d'utilisateurs de Facebook dans le monde arabe. Au cours du premier trimestre 2011, ce nombre fit un bond de 30 %, avec 27 711 503 utilisateurs recensés. Les pays membres du Conseil de coopération du Golfe (GCC) occupent les cinq premières places du classement du nombre d'utilisateurs arabes de Facebook en pourcentage de leur population. L'Égypte représente environ un quart du total des utilisateurs de Facebook dans les pays arabes, et a connu durant le premier trimestre 2011 la plus forte progression de nouveaux inscrits dans la région, soit 2 millions de nouveaux utilisateurs entre le 5 janvier 2011 et le 5 avril de la même année⁴.

Les gouvernements profitent de l'opportunité offerte par la gigantesque concentration de leurs citoyens sur Facebook pour renforcer leur présence dans les médias sociaux. L'Égypte et la Tunisie en sont des exemples frappants : immédiatement après le renversement des présidents par les mouvements de révolte, les gouvernements de transition se sont adressées aux Tunisiens et aux Égyptiens par le biais de Facebook. Bien que chaque gouvernement disposa déjà d'une présence non négligeable sur internet, et que chaque ministère et organisation nationale ait son propre site web, les nouveaux gouvernements créèrent des profils sur Facebook et YouTube, et ouvrirent des comptes Twitter, pour transmettre leur message et tenter d'engager une discussion avec leurs électeurs, que ces derniers résident dans leur pays ou bien à l'étranger. De manière analogue, le gouvernement des Émirats Arabes Unis (EAU), qui n'était à cette période confronté à

2 <http://www.dsg.ae/NEWSANDEVENTS/UpcomingEvents/ASMRHome.aspx>

3 <http://www.dsg.ae/NEWSANDEVENTS/UpcomingEvents/ASMROverview2.aspx>

4 <http://www.dsg.ae/portals/0/ASMR2.pdf>, page 9.

aucun conflit, ressentit la nécessité de se créer un profil et une présence sur Facebook afin de se rapprocher des 45 % de sa population connectés à ce réseau. Le gouvernement des EAU encourage désormais ses employés à utiliser les médias sociaux pour interagir avec les citoyens. Il a même proposé des formations sur ce qu'est un usage responsable de Facebook et les risques liés, et a adressé des directives d'utilisation détaillées à l'ensemble des organes gouvernementaux.⁵

DES USAGES VARIÉS

Au regard de la démographie des pays Arabes, où 30 % de la population est âgée entre 19 et 25 ans, et de leurs situations politique et économique, Facebook « est utilisé d'un grand nombre de façons différentes: rassembler des gens autour de revendications sociales et de campagnes politiques, améliorer le journalisme citoyen et la participation civique, créer un forum où les gouvernements et les habitants peuvent débattre et interagir, ou encore favoriser l'innovation et la collaboration au sein même du gouvernement »⁶. Cependant, l'usage principal de Facebook fut et demeure ce pour quoi il avait été conçu par ses créateurs: le réseautage social entre individus et groupes d'amis. Malgré la censure et les blocages, Facebook reste l'outil de mise en réseau par excellence pour les populations jeunes qui souhaitent communiquer, se rencontrer, partager leurs passions et leurs rêves, et suivre la vie de célébrités.

Au cours du premier trimestre 2011, le rôle du cyber-activisme dans les révoltes battant le pavé a déclenché une évolution radicale de l'usage et de la perception de Facebook dans les pays Arabes. Bloquer Facebook et l'internet en Tunisie, en Égypte, en Libye et en Syrie n'a fait que renforcer la confiance en ces outils et leur impact, provoquant une demande et une utilisation accrues.

YouTube et les réseaux de vidéo similaires sont utilisés depuis janvier 2011 en Égypte pour documenter et partager des événements allant d'appels à réunions publiques et manifestations jusqu'aux agressions et massacres perpétrés contre les manifestants. En Syrie, en Libye, et à Bahreïn, des vidéos datées ont permis à des protestataires de prouver la nature pacifique de leurs actions et la réponse brutale donnée par les

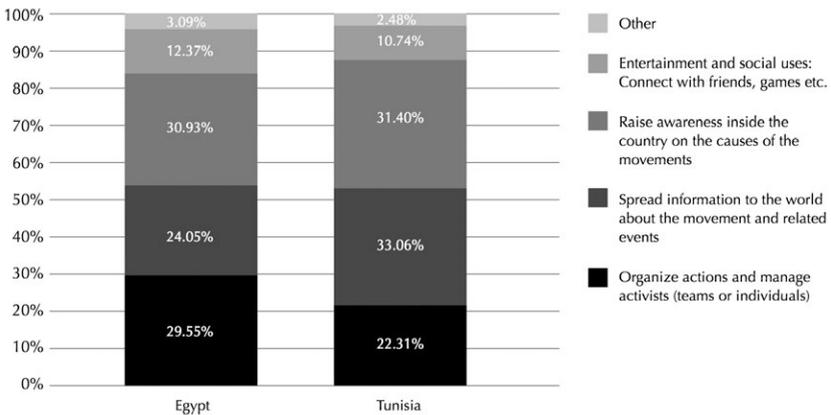
5 Guides by the government of UAE in arabic, <http://www.emiratesegov.ae/web/guest/83>

6 <http://www.dsg.ae/portals/0/ASMR2.pdf>, page 1.

autorités. Twitter fournit pour sa part un canal de communication pour des appels et rappels à rassemblement, des sos, des instructions rapides ou encore pour diffuser et recevoir les informations de dernière minute. La structure même de Twitter, dont les messages courts ou entrées de format « microblog » sont parfaitement adaptés aux téléphones mobiles, font de lui un dispositif de communication de masse particulièrement puissant.

Les rédacteurs du Rapport sur les médias sociaux dans le monde arabe ont interrogé 126 personnes en Égypte et 105 en Tunisie à propos de leur usage principal de Facebook pendant la révolte civile. Comme le montre le tableau ci-dessous :

Dans les deux pays, les utilisateurs de Facebook estimaient que ce dernier avait été principalement utilisé pour sensibiliser à l'intérieur même de leur pays aux mouvements civils en cours (31% en Tunisie et en Égypte), pour diffuser à l'étranger de l'information sur les mouvements (33% en Tunisie et 24% en Égypte), et pour organiser les actions et les activistes (22% en Tunisie et 30% en Égypte). Moins de 15% des personnes interrogées dans les deux pays pensaient que Facebook était utilisé en premier lieu pour se divertir ou interagir avec ses connaissances⁷.



⁷ <http://www.dsg.ae/portals/0/ASMR2.pdf>, page 6.

LA LANGUE DE FACEBOOK

Facebook propose son interface dans plusieurs dizaines de langues, le site ayant bénéficié dans la majorité des cas de localisations effectuées par les utilisateurs eux mêmes⁸. Les personnes interrogées dans les pays arabes lors de l'enquête menée par le rapport sur les médias sociaux « *différent quant à leurs préférences concernant la langue d'interface* »⁹, les trois premières langues étant sans trop de surprise l'arabe, le français et l'anglais. Le sondage révèle une nette préférence pour l'anglais dans les pays du Golfe, à l'exception de l'Arabie Saoudite, et une nette préférence pour le français dans les pays du Maghreb et dans l'Union des Comores. Il est intéressant de prêter une attention particulière à l'Égypte et à la Tunisie, en raison des changements opérés au cours même de la période de révolte. En ce qui concerne leurs préférences de langues d'interface, en mai 2011 les utilisateurs en Égypte se répartissent de manière égale entre personnes privilégiant l'arabe (49.88 %) et celles choisissant des commandes en anglais (48.98 %) (soit des résultats similaires à ceux observés en Jordanie, en Libye et en Irak). Les utilisateurs Tunisiens montraient quant à eux une préférence sans appel pour l'interface en français (94.60 %), venant ensuite le choix d'une interface en anglais (2.72 %) et un très faible choix de l'arabe (1.56 %).

Quoi qu'il en soit, les préférences de paramétrage linguistique de l'interface ne nous renseignent pas nécessairement sur les langues d'échange de contenu entre utilisateurs¹⁰. Le HTML et l'Unicode permettent désormais aux navigateurs d'afficher des textes dans quasiment toutes les langues. « *Les Facebookers pratiquent une diversité [de langues] qui remettent en question les notions conventionnelles de multilinguisme comme combinaison de deux ou plusieurs monolinguisms* »

UNE RÉVOLUTION DE LANGUE RÉGIONALE

L'importance du rôle joué par la langue dans les soulèvements sociaux et politiques en Tunisie et en Égypte, ainsi que dans les autres pays du MENA

8 Voir dans cet ouvrage : Dwayne Bailey, *Localisation des logiciels : open source et multilinguisme numérique*.

9 <http://www.dsg.ae/portals/0/ASMR2.pdf>, page 14.

10 Ingrid Piller <http://www.languageonthemove.com/language-globalization/multilingualism-2-0>

n'est pas une surprise. Certains slogans scandés par les protestataires sont devenus des chansons et des symboles internationaux. « *Ben Ali, dégage!* » en Tunisie, ou bien *al-sh'ab yureed asqat al-nitham*, (« Le peuple réclame la chute du régime »), répété à Tunis, au Caire, à Damas, à Benghazi, et Sana'a, sont intrinsèquement liés aux révoltes. Les banderoles déployées par les manifestants au Caire étaient majoritairement écrites en arabe, mais certaines l'étaient en anglais, en français et même en hébreu.¹¹

Sur le front des médias sociaux, les soulèvements ont influencé positivement la créativité linguistique. Sans doute la nécessité de toucher un plus large public sur des questions brûlantes incitait-elle à utiliser la langue locale, augmentant ainsi la quantité de contenu en langue arabe publiée en ligne, tant sur les médias sociaux que sur des sites web traditionnels.

Cette hypothèse prend appui sur mes observations concernant 1) le nombre de nouveaux sites internet publiés en arabe par des journaux, des mouvements sociaux et des entités gouvernementales et 2) le nombre d'utilisateurs de médias sociaux qui écrivent désormais en arabe.

DES STATISTIQUES ÉLOQUENTES

La seconde édition du *Rapport sur les médias sociaux dans le monde arabe*, publiée en Mai 2011, renforce également cette hypothèse. Si nous comparons la langue d'interface et la langue utilisée par les Égyptiens et les Tunisiens pour communiquer durant le mouvement civil du premier trimestre 2011, nous pouvons observer une grande différence. La distribution des langues d'interface se fait comme suit¹²:

Pays	Arabe (% d'utilisateurs FB)	Anglais (% d'utilisateurs FB)	Français (% d'utilisateurs FB)
Égypte	49.88	48.98	0.39
Tunisie	1.56	2.72	94.60

¹¹ Voir l'image de gcarlstrom (Gregg Carlstrom) à <http://yfrog.com/h3fbsbj> traduction : *azov Mubarak* - "leave, Mubarak".

¹² Arab Social Media Report, Issue 2, May 2011, <http://www.dsg.ae/portals/0/ASMR2.pdf> page 14.

Cependant, lorsqu'on les interroge¹³ sur la langue principale utilisée pour communiquer sur Facebook durant les soulèvements populaires, les participants à l'enquête ont donné les réponses suivantes :

Pays	Arabe (% d'utilisateurs FB)	Anglais (% d'utilisateurs FB)	Français (% d'utilisateurs FB)
Égypte	75.40	25.60	0
Tunisie	51.43	0.95	47.62

Au regard des résultats de cette enquête, les révolutions en Égypte et en Tunisie sont la source de cette disparité accrue entre langue d'interface et langue de communication.

Pour mettre cette hypothèse à l'épreuve, j'ai procédé à l'examen des Tweets du 6 juillet 2011 qui étaient liés au mot-clé, ou *hashtag*, #CairoExplosion. Le 6 juillet 2011, les habitants du Caire et de sa banlieue entendirent une énorme explosion dont l'origine demeura inconnue de nombreuses heures durant. Rapidement, un mot-clé fut créé sur Twitter et ses utilisateurs commencèrent à envoyer des messages interrogeant, répétant, devinant ou renvoyant par tweets ce qu'ils avaient entendu. En moins d'une heure, ce *hashtag* généra suffisamment d'attention pour susciter un certain niveau de panique parmi les utilisateurs égyptiens de Twitter. J'ai sélectionné aléatoirement 500 tweets contenant le mot-clé #CairoExplosion, tous envoyés dans un intervalle d'environ deux heures, et les ai classés selon le critère de la langue utilisée, ce qui a généré la répartition suivante :

Arabe	Arabe en alphabet latin	Anglais	Mix des deux langues	Indéterminée	Total
353	33	89	19	6 (smileys)	500

Malgré l'intérêt évident de telles données, l'observation est limitée et mériterait d'être systématiquement examinée et validée. Certaines initiatives comme R-Shief¹⁴ visent, entre autre choses, à collecter et visualiser des contenus de Twitter et de la sphère publique de Facebook pour étudier la distribution des langues et de leurs usages.

Cependant, augmenter la quantité de contenu ne signifie pas nécessairement améliorer sa qualité. Les tweets sont limités à de courts messages et à des entrées de microblogs. Le contenu des médias sociaux est caractérisé

¹³ Arab Social Media Report, Issue 2, May 2011, <http://www.dsg.ae/portals/0/ASMR2.pdf> page 7.

¹⁴ <http://www.r-shief.org/>

par son style informel, mais il est souvent repris par des réseaux et des pages internet plus institutionnels, qui, en sus de la présence active d'entités et d'organisations gouvernementales, lui confèrent dès lors un statut plus formel. Un grand nombre d'organisations internationales, dont le Programme des Nations Unies pour le Développement (PNUD), le Fond International pour le Développement de l'Agriculture (FIDA) et le Centre Canadien de Recherche sur le Développement International (CRDI/IDRC) sont actifs sur leurs pages Facebook et leur compte Twitter, contribuant à améliorer le style et la crédibilité de ces outils. Et depuis la campagne présidentielle de Barack Obama en 2008, presque tous les candidats politiques des pays développés ou en voie de développement utilisent les médias sociaux pour communiquer avec leurs électeurs, ou du moins la partie utilisant Facebook.

CONCLUSION

Partout dans le monde apparaissent clairement des signes d'une préférence pour les médias sociaux comme outils de communication, de mobilisation, de manifestation et d'expression du mécontentement de la population, lors de soulèvements ou de période d'agitation sociale. Les pays du Moyen Orient et d'Afrique du Nord ont ainsi vu le nombre d'utilisateurs des médias sociaux augmenter sensiblement. Ces utilisateurs communiquent essentiellement dans leur langue première, quelle que soit leur préférence au regard de la langue d'interface. Une recherche plus approfondie est nécessaire pour mieux documenter ce phénomène et développer les hypothèses présentées ici afin d'en améliorer la qualité. Mais le plus important est qu'une telle recherche pourrait contribuer à améliorer la participation citoyenne via le cyberspace, et rendre patents les bénéfices grandissants générés par la société de la connaissance.

MULTILINGUISME, OBJECTIFS DU MILLENAIRE POUR LE DÉVELOPPEMENT ET LE CYBERESPACE

Les « Objectifs du Millénaire pour le Développement » (OMD), définis par les Nations Unies en 2000, doivent être évalués lors d'un Sommet mondial en 2013. Comment l'usage des langues dans les technologies peut constituer une aide déterminante pour leur réalisation ? Comment les politiques linguistiques publiques, locales ou multilatérales, peuvent participer à atteindre ces objectifs partagés ?

Article original en français.



ADAMA SAMASSÉKOU est Président du Conseil International de la Philosophie et des Sciences Humaines (CIPSH) et Président du Réseau MAAYA, Réseau Mondial pour la Diversité Linguistique. Ancien Secrétaire Exécutif de l'Académie Africaine des Langues (ACALAN), Institution spécialisée de l'Union Africaine basée à Bamako (Mali), et Président du Comité Préparatoire (PrepCom) du Sommet Mondial sur la Société de l'Information pour la phase de Genève (SMSI/ 2002-2003), il a été Ministre de l'Education de Base du Mali, initiateur de la Refondation du Système éducatif (1993-2000) et Porte-parole du Gouvernement (1997-2000). Membre du Haut Conseil de la Francophonie de 2003 à 2006, il est aujourd'hui membre de la Commission Internationale de l'IUT et de l'Unesco de la Bande Passante pour le Développement Numérique.

ADAMA SAMASSÉKOU

MULTILINGUISME,
OBJECTIFS DU
MILLENAIRE POUR
LE DEVELOPPEMENT
ET LE CYBERESPACE

Lorsqu'on parle de multilinguisme, on pense souvent à la performance linguistique, à la façon qu'ont les hommes de communiquer, d'enregistrer leurs cultures et de les confronter ou les partager. Lier la question de la langue et les questions dites de développement pourrait sembler peu pertinent, tellement c'est un lien naturel et de bon sens qui n'a point besoin d'être mis en exergue ! Seulement, il y a tellement de pays dans le monde où ce lien est rendu complexe par l'histoire, par les effets à long terme de la colonisation, au point que la relation entre les langues et le développement est biaisée, voire niée, qu'il devient fondamental de se pencher sur lui et lui donner l'éclairage requis. Alors que nous entrons dans une nouvelle société en construction rapide, celle de l'information, de la connaissance et des savoirs partagés, l'apport du multilinguisme aux Objectifs du Millénaire pour le Développement (OMD) mérite qu'on prenne le temps de l'analyse. Le cyberspace prend une place majeure dans la réorganisation du monde et les projets collectifs. Les orientations des organisations multilatérales ont aussi des impacts majeurs sur les relations virtuelles. Le multilinguisme n'échappe pas à cette focalisation de l'avenir du globe sur les technologies de la communication et de l'information.

ÉLIMINER LA PAUVRETÉ

Les huit Objectifs du Millénaire pour le Développement (OMD), ont été adoptés lors du Sommet du Millénaire qui s'est déroulé du 6 au 8 septembre 2000, au Siège des Nations Unies à New York.

« L'élimination de l'extrême pauvreté demeure l'un des grands défis de notre temps et constitue l'une des principales préoccupations de la communauté internationale. [...] Les Objectifs du Millénaire pour le Développement sont assortis de cibles à atteindre dans des délais

précis, de façon à mesurer les progrès accomplis dans la lutte contre la pauvreté monétaire, la faim, la maladie, l'absence de logements adéquats et l'exclusion tout en promouvant l'égalité des sexes, la santé, l'éducation et le respect de l'environnement. [...] Ces objectifs ambitieux mais réalisables indiquent la voie à suivre par la communauté internationale pour faire reculer l'extrême pauvreté d'ici à 2015 dans le cadre du programme global de l'ONU pour le développement»¹.

Les 8 objectifs visant à une amélioration notable des conditions de vie sur terre avant 2015 sont les suivants : 1. réduire l'extrême pauvreté et la faim ; 2. assurer l'éducation primaire pour tous ; 3. promouvoir l'égalité des sexes et l'autonomisation des femmes ; 4. réduire la mortalité infantile ; 5. améliorer la santé maternelle ; 6. combattre le VIH/SIDA, le paludisme et d'autres maladies ; 7. assurer un environnement humain durable ; 8. mettre en place un partenariat mondial pour le développement.

Alors que l'année 2015 approche, nous constatons aisément que ces Objectifs restent en chantier, même si pour certains d'entre eux nous pouvons apercevoir des améliorations notables. De ce point de vue, l'initiative prise récemment par l'Union Internationale des Télécommunications (UIT) et l'Unesco de créer une *Commission Internationale de la Bande Passante pour le Développement numérique*, en vue d'atteindre plus rapidement les Objectifs du Millénaire pour le Développement, mérite d'être signalée². L'évolution du monde conduit aujourd'hui à intégrer les OMD dans la nouvelle négociation internationale sur le développement durable qui se déroulera à Rio en juin 2012. Dans celle-ci les questions d'environnement et de développement durable, d'équilibre entre les pays et de protection de la planète face aux défis climatiques et l'érosion de la biodiversité seront mises en regard de l'éradication de la pauvreté approuvée par l'ensemble des pays en 2000. Si le lien avec un économie plus sensible aux questions environnementales semble évident, l'approche par l'expansion de la culture et le partage de la connaissance reste trop peu élucidée.

Le développement est une notion complexe, qui renvoie à l'histoire, et aux inégalités dans le monde qu'elle nous a laissé. L'analyse présentée dans

1 M. Ban ki-moon, Secrétaire général des Nations Unies, *Passer à l'action pour réaliser les objectifs du Millénaire pour le développement*, 25 juillet 2008 <http://www.un.org/millenniumgoals/2008highlevel/pdf/0840769f.pdf>

2 <http://www.broadbandcommission.org>

cet article concerne donc uniquement certains pays du monde – ceux pour lesquels la Communauté internationale a identifié les Objectifs de Développement pour le Millénaire, à atteindre à l’horizon 2015.

Pour la clarté de la démonstration, nous avons fait le choix de nous concentrer sur le cas de l’Afrique, où la problématique du lien entre multilinguisme, OMD et cyberspace est la plus pertinente, parce que la plus urgente à prendre en compte. Et d’insister ici sur la dimension linguistique des objectifs et des moyens de les atteindre.

Il est important de souligner la spécificité de la question linguistique dans les pays anciennement colonisés en général et en Afrique en particulier, qui reste marquée par une quête identitaire permanente dont les origines sont bien connues.

La *Charte culturelle de l’Afrique*³, adoptée par les Chefs d’État et de Gouvernement de l’Organisation de l’Unité Africaine (OUA), réunis en leur Treizième Session Ordinaire à Port Louis (Île Maurice) du 2 au 5 juillet 1976, nous rappelle « *que sous la domination coloniale, les pays africains se sont trouvés dans une situation politique, économique, sociale et culturelle identique; que la domination sur le plan culturel, a entraîné la dépersonnalisation d’une partie des peuples africains, falsifié leur histoire, systématiquement dénigré et combattu leurs valeurs africaines, tenté de remplacer progressivement et officiellement leurs langues par celle du colonisateur; que la colonisation a favorisé la formation d’une élite trop souvent acculturée et acquise à l’assimilation et qu’une grave rupture s’est produite entre cette élite et les masses populaires africaines* ».

La gestion appropriée d’une telle situation exige donc pour ces pays de faire un choix de projet de société qui soit fondé sur un développement endogène et qui garantisse une diversité culturelle et linguistique permettant de faire en sorte que toutes les cultures et toutes les langues puissent s’exprimer et se faire connaître. Ceci sera réellement possible quand les langues africaines pourront être présentes comme instruments de travail dans tous les domaines de la vie publique, en partenariat avec les langues officielles héritées de la colonisation, dans tous les moyens d’expression, de communication et de diffusion.

3 Charte culturelle de l’Afrique http://www.africa-union.org/Official_documents/Treaties_Conventions_fr/Charte%20Culturelle.pdf

D'où la nécessité et l'urgence de concevoir une perspective africaine d'aménagement linguistique. On le sait, l'aménagement linguistique consiste à prendre des mesures politiques et techniques pour préserver ou promouvoir l'équilibre linguistique d'un territoire où sont parlées plusieurs langues, afin de mieux servir un projet de société.

L'adoption, pour tous ces pays anciennement colonisés, d'une politique linguistique explicite reflétant le projet de société choisi s'avère ainsi une option incontournable.

LE MULTILINGUISME, UNE RÉALITÉ VÉCUE DANS LES PAYS CONCERNÉS PAR LES OMD

Si, pendant les premières années des indépendances africaines, la culture guidait et éclairait les choix politiques parce que considérée comme la base et la finalité de tout processus de développement social et économique, force est de reconnaître qu'aujourd'hui ce sont, malheureusement, les programmes de développement exogènes qui inspirent de plus en plus les projets de promotion des valeurs culturelles. C'est pourquoi nous pensons qu'il est urgent d'inverser cette tendance, parce qu'à notre avis il est utopique de concevoir des projets de développement en dehors des valeurs culturelles des populations auxquelles ces projets sont destinés.

Cependant, dans la grande majorité des États concernés, le paradoxe de la situation réside principalement dans le fait que nous sommes en présence d'États plurilingues mais qui, héritiers de la tradition coloniale, se trouvent face à un choix pratiquement imposé de l'État-nation privilégiant la perspective monolingue.

C'est pourquoi, pour éviter ce choix contraire à la réalité de ces pays, nous prônons la mise en œuvre d'une approche stratégique que nous avons appelée multilinguisme fonctionnel convivial et que nous définissons ainsi : *« une approche stratégique de gestion du pluralisme linguistique africain prenant en compte à la fois le principe d'équité des langues et le principe de réalité des différentes fonctions assumées par celles-ci. Cette approche s'articule avec la décentralisation administrative et l'intégration africaine et a comme matrice une langue identitaire, une langue véhiculaire et une langue de communication internationale. Elle prône la convivialité des langues et la « délégation de souveraineté linguistique » par la pratique du principe de subsidiarité entre les niveaux local, régional, national et africain ».*

LE MULTILINGUISME, CONDITION SINE QUA NON DE LA RÉALISATION DES OMD ET D'UN CYBERESPACE INCLUSIF

La très grande majorité de l'humanité vivant dans des sociétés plurilingues, où le multilinguisme est la norme, comment donc ne pas prendre en compte la question linguistique, le multilinguisme dans la réalisation des OMD ? Question transversale par excellence, la question de la langue conditionne l'atteinte et la réalisation de chacun des huit Objectifs du Millénaire pour le Développement. Essayons de mesurer cet impact pour chacun d'entre eux.

Réduction de l'extrême pauvreté et de la faim dans le monde

Il n'y aura ni réduction, encore moins éradication de la pauvreté, tant que les acteurs à la base n'auront pas été capacités dans leur(s) langue(s), qui devraient être leurs instruments ordinaires de travail, d'appréhension et de transformation de leur environnement, au sens large du terme. La faim dans le monde sera réduite en mobilisant les communautés paysannes. Or celles-ci parlent des langues du terroir, qui décrivent au mieux leur terre, les conditions de la culture, les fruits et légumes adaptés aux territoires. Le groupe d'experts international « Évaluation Internationale des Connaissances, des Sciences et des Technologies Agricoles pour le Développement »⁴ dans son rapport de 2008 indique que près d'un milliard d'humains vont se coucher sans avoir mangé à leur faim. Il estime que le meilleur moyen de lutter contre cet état de fait est de favoriser les petites exploitations, et d'apporter aux communautés rurales les connaissances, la technologie et le crédit⁵. Tout ces savoirs doivent être exprimés dans des langues que les communautés paysannes peuvent comprendre, assimiler et enrichir de leurs savoirs propres. On voit ici que la question linguistique est en impact direct avec la réduction de la faim dans le monde.

4 EICSTAD, Évaluation Internationale des Connaissances, des Sciences et des Technologies Agricoles pour le Développement / IAASTD, International Assessment of Agricultural Knowledge, Science and Technology for Development <http://www.agassessment.org>

5 Global food system 'must change, Mark Kinver, *BBC News*, 15 avril 2008 <http://news.bbc.co.uk/2/hi/science/nature/7347239.stm>

Assurer l'éducation primaire pour tous

Il n'y aura pas d'instruction primaire universelle de qualité tant que ne se mettra pas en place un enseignement bi- et multilingue fondé sur la langue maternelle⁶. On commence à entrevoir les capacités du cyberspace pour étendre l'éducation pour tous. Mais un grand risque existe que les manuels, les exemples, les modèles ne soient conçus et rédigés uniquement dans les langues dominantes et pour les cultures des pays les plus développés. Accéder aux concepts scientifiques dans sa langue maternelle, puis les traduire ultérieurement dans la scolarité dans d'autres langues permet une meilleure appropriation et un véritable échange de savoirs.

La formation des maîtres, qui doivent souvent maîtriser plusieurs langues pour suivre les cours des universités comme pour interagir avec leurs élèves peut largement s'appuyer sur les TIC. De même que l'échange de cours et de ressources éducatives. Les Ressources éducatives libres⁷, qui autorisent la traduction et la réutilisation dans d'autres contextes forment un outil puissant pour réaliser cet objectif tout en s'appuyant sur un cyberspace multilingue.

Promouvoir l'égalité des sexes et l'autonomisation des femmes

Les indicateurs choisis par l'ONU pour rendre compte des progrès en la matière sont les suivants⁸:

- le rapport filles/garçons dans l'enseignement primaire, secondaire et supérieur;
- le taux d'alphabétisation des femmes de 15 à 24 ans par rapport aux hommes;
- le pourcentage de femmes salariées dans le secteur non agricole;
- la proportion de sièges occupés par des femmes au parlement national.

6 Voir dans cet ouvrage: Marcel Diki-Kidiri, *Le cyberspace dans l'éducation en langue maternelle*.

7 *Capeton Education declaration*, <http://www.capetowndeclaration.org>

Une traduction en français est disponible: Déclaration du Cap: débrider le potentiel des ressources éducatives partagées, In: *Libres Savoirs, les biens communs de la connaissance*, C&F éditions, 2010, p. 169-173.

8 Déclaration du Millénaire et OMD, In: *Égalité des hommes et des femmes en Francophonie*, <http://genre.francophonie.org/spip.php?article384>

Pour l'ensemble de ces critères, la question du nombre de langues maîtrisées apparaît centrale. Même si les résultats sur ces critères sont loin d'être à la hauteur des espoirs, l'éducation des femmes a largement progressé. Mais de nombreux obstacles s'installent dès qu'il s'agit d'une éducation au-delà du niveau primaire. Dans ce cadre, l'enseignement à distance peut offrir un outil. Mais pour autant, les formations doivent être dispensées dans les langues des apprenants et offrir des possibilités de traduction.

Une attention particulière doit être portée à l'apprentissage des technologies de l'information et de la communication par les femmes, soit dans l'éducation formelle, soit au sein de groupes de femmes orientés vers la technologie. Pour favoriser ce travail collectif, les femmes doivent pouvoir échanger dans leur propre langue, et savoir utiliser les outils soit dans leur langue, soit dans une des langues véhiculaires de leur région. Or la plupart des logiciels et services se présentent dans les quelques langues dites majeures, ce qui crée des obstacles supplémentaires. Un effort de localisation des logiciels⁹ va renforcer les capacités des TIC et du cyberspace à appuyer la réalisation de cet objectif central pour l'égalité dans le monde.

Réduire la mortalité infantile ; Améliorer la santé maternelle

Ces deux objectifs restent intimement liés. Les sociétés devraient faire de l'accueil des enfants sur la terre l'une de leur raison d'être. Or les femmes des pays en développement sont en grand risque lors de leur grossesse et des accouchements. Un enfant né dans un pays en développement a 13 fois plus de risques de mourir dans ses 5 premières années de vie qu'un enfant né dans un pays industrialisé.

Là encore, de nombreux progrès ont été réalisés, notamment dans la présence de personnel soignant lors des accouchements, et dans la prévention de maladies que l'on sait guérir et qui sont particulièrement violentes pour les jeunes enfants (pneumonies, diarrhées, rougeole).

La réduction de la mortalité infantile et maternelle est intimement liée à la capacité pour les enfants et les mères de prendre conscience de la

⁹ Voir dans cet ouvrage : Dwayne Bailey, *Localisation des logiciels : open source et multilinguisme numérique*.

pertinence des messages de sensibilisation sanitaire conçus dans leurs langues ; à la communauté de reprendre confiance dans les savoirs, savoir-être et savoir-faire endogènes multiséculaires de son peuple tels qu'ils sont incrustés dans sa et ses langue(s).

L'usage du multimédia⁴⁰ pour favoriser cette diffusion des messages de santé publique adaptés aux conditions réelles des femmes et des enfants des villes et villages les plus pauvres peut être améliorée. Notamment le doublage et l'usage des langues nationales.

Combattre le VIH/SIDA, le paludisme et les autres maladies

Il en est de même en ce qui concerne l'arrêt et l'inversion des tendances de diffusion du VIH/SIDA, du paludisme et des autres maladies endémiques. Malgré les progrès de la médecine et des antirétroviraux, le SIDA reste une plaie pour les pays en développement. Les messages de prévention, l'utilisation de méthodes contraceptives, l'éducation sexuelle demeurent des moyens essentiels pour contrer l'extension de cette maladie. Le rôle des diffusions culturelles dans cette formation des mentalités à une échelle jamais égalée jusqu'à présent, est rendu encore plus facile avec le cyberespace. Interconnecter les radios communautaires rend plus aisé l'échange de déclarations, d'interview, de spots musicaux... par des représentants des scènes musicales ou culturelles qui incitent à la prévention. En Afrique, la radio reste le moyen le plus employé pour communiquer des informations, mais l'internet est un outil pour construire des réseaux de radios communautaires si on prend soin d'y développer le multilinguisme.

Pour l'ensemble des maladies infectieuses, le cyberespace, parce qu'il permet de diffuser les dernières recherches de pointe en matière médicale est un outil très performant. Les équipes médicales des pays en développement doivent pouvoir accéder aux dernières recherches et aux résultats des expériences. C'est toute la chaîne de la santé, depuis les universités jusqu'aux dispensaires qui est concernée par le partage des savoirs médicaux. Le libre accès aux publications médicales⁴¹ et les

10 Voir dans cet ouvrage : Tunde Adegbola, *Multimédia, langue des signes, langues écrites et orales*.

11 Leslie Chan, Subbiah Arunachalam & Barbara Kirsop, La chaîne de la communication dans les sciences de la santé : des chercheurs aux praticiens, l'impact du libre accès, In : *Libres savoirs, les biens communs de la connaissance*, C&F éditions, 2010, p. 86-104.

traductions des résultats essentiels dans le plus grand nombre de langues font du cyberspace multilingue un soutien aux pratiques médicales dans le monde entier, et notamment dans les lieux les plus éloignés des centres de recherche et des hôpitaux de pointe.

D'une manière générale, les progrès de la télémédecine dans un pays comme le Mali ouvrent d'excellentes perspectives de prise en charge à distance de la maladie et des malades plus à l'aise pour expliquer leur maladie dans leurs langues, en particulier pour les zones rurales et peu dotées d'infrastructures médicales

Préserver l'environnement

L'environnement est le produit d'un rapport culturel à la nature. Chaque société participe à la protection de la biodiversité, et malheureusement parfois à sa destruction, par son action physique sur le monde qui l'entoure. Mais elle le fait aussi par les mythes et légendes, les écrits et les recherches qui façonnent l'esprit des citoyens. Le recours aux pratiques ancestrales, liées à une cosmovision fondée sur l'harmonie et le respect de la nature et qui ont traversé les siècles grâce à la tradition orale soutenue par les langues, est plus à même de rendre les questions environnementales plus familières et rapidement appropriables.

Parce qu'il permet de faire cohabiter ces visions du monde, qu'il revivifie chez les urbains le long fil qui les relie à la nature, le cyberspace devient un outil pour que les sociétés prennent conscience des dangers que représentent le changement climatique global et l'érosion de la biodiversité.

C'est aussi parce qu'il permet de documenter les atteintes à l'environnement, de faire savoir au-delà des lieux directement concernés, que le cyberspace multilingue renforce la conscience globale de l'environnement.

Mettre en place un partenariat pour le développement

Comment développer des partenariats multi-acteurs féconds au service du développement, au niveau mondial et local, sans partager les mêmes langues de communication ?

La nouvelle société, celle de la connaissance et des savoirs partagés, qui se construit à travers les TIC, est désormais en marche... Comment faire en sorte qu'elle soit totalement inclusive et qu'aucune couche, aucune communauté, aucune société ne soit laissée pour compte?

CONCLUSION

Les sociétés concernées au premier chef par les OMD sont en même temps celles qui ne sont pas encore entrées dans la culture du numérique, essentiellement parce que le fossé numérique se double d'un fossé linguistique, leurs langues n'ayant pas encore droit de cité dans l'internet...

Comment garantir à chaque langue la possibilité d'être utilisée dans le cyberspace et dans l'internet comme moyen de communication, de production et de partage de l'information et des savoirs, conformément aux recommandations du Sommet mondial sur la Société de l'Information (SMSI)?

Au regard de la transversalité de la question linguistique pour la réalisation des OMD, il apparaît finalement indispensable d'intégrer le multilinguisme comme un nouvel Objectif du Millénaire

- multilinguisme, facteur de capacitation des sociétés plurielles;
- multilinguisme, facteur de consolidation des identités plurielles et de renforcement du développement sociétal;
- multilinguisme, facteur de renforcement de la confiance en soi et de construction d'un vrai dialogue des cultures, gage de cohésion sociale, de paix et de compréhension mutuelle.

Si le multilinguisme permet d'accélérer le rythme d'approche de réalisation des OMD, alors il doit devenir lui-même le neuvième Objectif du Millénaire pour le Développement.



**LE MULTI-
LINGUISME
SUR
INTERNET,
UNE
QUESTION
MULTI-
LATÉRALE**

PARTIE 4

La question des langues dans le cyberspace est une question proprement politique à l'échelle multilatérale. De la définition des normes de gouvernance du réseau jusqu'à la relation entre la défense du multilinguisme et les Droits humains, les langues peuvent être convoqués comme des repères majeurs dans la société de l'information.

DÉCRIRE LE MONDE : MULTILINGUISME, INTERNET ET DROITS DE L'HOMME

Promotion et protection du multilinguisme sur internet participent-elles de l'architecture internationale des Droits humains ? Pour concevoir les enjeux juridiques du multilinguisme dans le cyberspace, il convient d'évaluer les liens pouvant exister entre la philosophie des lois, la théorie des actes de langage et les universaux transculturels en regard des usages de l'internet et du web. Que serait un droit linguistique humain ? Comment définir sa nature et sa portée ? Comment les changements des régimes juridiques, publics ou privés, locaux ou internationaux, ont des effets sur le cyberspace ?

*Celui qui a voulu que l'homme fût sociable
toucha du doigt l'axe du globe et l'inclina sur l'axe de l'univers.*
Jean-Jacques Rousseau
Essai sur l'origine des langues, Œuvres, Belin, Paris, 1817

Article original en anglais.
Traduit par Francisca Cabrera.



ISABELLA PIERANGELI BORLETTI est titulaire d'un doctorat en droit international et en droit européen, soutenu à l'université de Pise, Italie. Elle est actuellement chercheuse invitée à l'Ankara University European Research Center (ATAUM), Turquie, où elle enseigne le droit international public, le droit européen et le droit des organisations internationales.

ISABELLA PIERANGELI BORLETTI

L'HOMME
DROITS DE
INTERNET ET
MULTILINGUISME,
LE MONDE :
DÉCRIRE

Nous nous proposons dans cet article d'évaluer le potentiel des orientations juridiques nationales, régionales et internationales pour promouvoir et protéger le multilinguisme sur internet en référence au cadre des Droits de l'Homme. Une exploration en profondeur de cette question requiert la compréhension des liens existant entre le droit international public et privé, ainsi que des normes sociales et des universaux culturels associés à l'utilisation de l'internet dans le monde.

Le développement de plus en plus rapide de « l'immatériel en ligne » conduit à se poser un certain nombre de questions touchant aux Droits de l'Homme, dont celle du multilinguisme¹ comme élément intrinsèque de la diversité culturelle. Si le multilinguisme est considéré comme faisant partie des Droits de l'Homme, la protection des droits individuels et collectifs des minorités linguistiques dans le cyberspace est un moyen de garantir que d'autres droits humains, tel que les droits à l'éducation, au développement, à la liberté d'expression, ainsi que les droits de rechercher et d'accéder à l'information. Le défaut de reconnaissance pleine et entière du multilinguisme sur internet entrave l'exercice effectif de ces droits.

Une analyse de l'état de la réflexion contemporaine et des outils juridiques adoptés par les États et les organisations internationales, permet d'éclairer

1 Pour une plus grande clarté, soulignons ici la différence entre multilinguisme et plurilinguisme comme proposée par le Conseil de l'Europe dans son document sur la politique linguistique éducative : http://www.coe.int/t/dg4/linguistic/division_FR.asp. Le multilinguisme désigne la présence de plusieurs langues dans une aire géographique donnée, grande ou petite, et donc des formes verbales appartenant à des groupes sociaux, reconnues ou pas formellement comme des langues. À l'intérieur de cette aire les individus pourront être monolingues (ou unilingues) parlant seulement leur propre variante. Le plurilinguisme se rapporte au répertoire des langues utilisées par plusieurs individus. C'est donc l'opposé du monolinguisme, incluant une « langue maternelle » ou une « langue première » et bien d'autres variantes linguistiques. Dans les aires multilingues, certains individus sont monolingues et d'autres plurilingues.

jusqu'à quel point sont solidaires – ou, au contraire déconnectés, les trois concepts : Droits de l'Homme, Internet et multilinguisme ; afin de fournir une base pour des cadrages futurs.

INTERNET ET DROITS DE L'HOMME

Courriers électroniques, blogs, moteurs de recherche, commerce en ligne et podcasts ne mettent pas nécessairement en jeu tous les Droits de l'Homme, mais le rapide développement d'internet dans nos sociétés soulève de nombreuses questions au regard des droits civils, politiques, économiques et sociaux, tant dans leurs dimensions individuelles que collectives, appelant dès lors à un traitement transversal de la question.

Nous sommes témoins depuis quelque temps de l'émergence d'une corrélation entre l'internet et les Droits humains, et ceci particulièrement dans des contextes où la liberté d'expression et les droits au respect de la vie privée sont menacés par des politiques d'État de contrôle et de censure. La Loi turque n°5651 du 4 Mai 2007, limitant la liberté d'expression et réduisant l'accès des citoyens à l'information², est un exemple de la façon dont les politiques d'État peuvent, en limitant l'accès à internet, nuire à d'autres droits. Le langage juridique international définit les Droits de l'Homme comme « *universels et inaliénables, indivisibles et interdépendants* ».

La nécessité d'un cadre conceptuel fondé sur les droits, dans le débat sur l'accès à internet et son utilisation, est de plus en plus évidente. Cette approche devrait donc être défendue au plus haut niveau politique, et doublée d'une attention singulière à porter aux notions de développement durable, de fracture numérique, de multilinguisme et d'éthique dans le cyberspace.

2 Dans la partie de son Rapport sur la liberté des médias consacrée à la Loi sur l'Internet de la Turquie, publié le 11 janvier 2011, la Représentante de l'OSCE, Mme Miklos Harszti, incite fortement les autorités Turques à abolir ou amender cette Loi.

INTERNET COMME UN BIEN COMMUN

Internet et le droit au développement

Le droit au développement a été choisi comme thème central du Forum sur la Gouvernance de l'Internet³. C'est un concept qui fait l'objet de débats dans les milieux tant académiques et politiques. La reconnaissance comme « Droit de l'Homme » en tant que tel est en cours⁴ et l'obligation légale des pays développés d'apporter assistance aux pays en développement est en discussion au sein de l'Organisation des Nations Unies (ONU)⁵. D'un point de vue juridique, la notion de développement inclut quatre facteurs : *responsabilité, autonomisation, participation et non-discrimination*.

- *La responsabilité* fait référence à l'identification d'obligations spécifiques et définit les entités qui en sont responsables, faisant ainsi glisser la coopération au développement du domaine de la charité à celui du droit, impliquant des titulaires d'obligations dans la préservation de l'accès au bien commun ;
- *L'autonomisation* signifie que les activités dédiées au développement doivent faciliter et soutenir les efforts fournis par les communautés pour améliorer leurs conditions de vie et prendre en main leur destin ; Pouvoir, capacité et accès doivent pour cela être alloués à ces communautés. Par exemple, l'Observatoire Mondial sur la Société de l'Information⁶ (GISWatch) a récemment mené une étude sur les pêcheurs des côtes Sud de l'Inde, qui ont obtenu accès à l'information

3 L'objectif du Forum est de soutenir le Secrétaire général des Nations Unies dans l'exécution de son mandat attribué lors du Sommet Mondial sur la Société de l'Information (SMSI) et visant à l'organisation d'un nouveau forum ouvert à un dialogue entre les multiples parties prenantes sur la gouvernance de l'internet. Le Forum sur la Gouvernance de l'Internet a concentré ses questions sur l'objectif d'accroître le potentiel d'Internet à son maximum, et de la sorte stimuler le bien-être social et économique des pays en voie de développement.

4 A. Esterhuysen, R. Greenstein, « The Right to Development in the Information Society », in *Human Rights in the Global Information Society*, MIT Press, Boston, 2006, p. 285.

5 Les bases juridiques du droit au développement peuvent être trouvées dans les documents suivants : *Déclaration Universelle des Droits de l'Homme*, Articles 2(1) et 2(3) ; *Charte d'Afrique relative aux droits des peuples et de l'homme*, Article 2(1) ; la *Déclaration du droit au développement de l'ONU* (1986) ; la *Déclaration et programme d'action de Vienne* (1993), adoptée dans la Conférence Mondiale sur les droits de l'homme ; et le *Rapport de 2000 sur le développement humain* (UNDP) des Nations Unies.

6 <http://www.giswatch.org/about>

sur les conditions météorologiques et le marché de la pêche dans leur propre langue par le biais de leurs téléphones mobiles⁷ ;

- *La participation* désigne un degré élevé de participation des communautés, de la société civile, des groupes minoritaires, des populations autochtones, des femmes ou des autres groupes concernés ;
- *La non-discrimination et l'attention portée aux groupes vulnérables* soulignent la nécessité d'éviter le renforcement des asymétries pré-existantes en termes de distribution du pouvoir et des ressources, en donnant la priorité aux groupes défavorisés.

Ces principes devraient être analysés à la lumière des thèmes principaux développés par le Forum sur la Gouvernance de l'internet. L'objectif serait dès lors d'optimiser le potentiel propre à l'internet d'amélioration du bien-être social et économique du plus grand nombre de personnes possibles dans les pays en développement, suivant les principes d'accès à l'information, diversité, ouverture et sécurité. Parmi ces quatre principes, la question de l'accès est d'une importance particulière dans la mesure où elle est directement liée au droit au développement.

En termes pratiques, cela veut dire que tous les individus devraient être à même de participer au progrès et aux bénéfices apportés par le développement technologique (selon les Articles 26 et 27 de la Déclaration Universelle des Droits de l'Homme) et que les pays en développement devraient avoir accès à internet en tant qu'outil primordial de développement de l'éducation et de l'économie. Cet objectif, qui ne peut être atteint que dans le renforcement des infrastructures, des équipements et compétences tournées vers l'universalisation de l'accès au Web, a été défendu sous différentes bannières telles que le « haut débit pour tous », le « service universel » ou encore les « services à caractère publics ».

Nombre d'acteurs nationaux et internationaux considèrent qu'une large couverture en haut débit est cruciale pour le soutien de la croissance économique, principalement dans les régions les moins développées. La reconnaissance par l'Europe de son obligation à permettre l'accès à Internet à tous les citoyens ne fait que souligner la disparité existant avec les pays en développement. Le rapport élaboré en 2010 sur les Objectifs du Millénaire pour le Développement (OMD) – l'objectif 8 particulièrement, *Développer un partenariat mondial pour le développement* – rend

⁷ S. Arunachalam, « Information and Livelihoods », in *Global Information Society Watch 2009 – Advancing Human Rights and Democracy*, disponible en ligne <http://www.giswatch.org>

manifestes les écarts existants en termes d'accès à Internet, ou encore la « fracture numérique » entre pays développés et pays en développement⁸. En 2008, 23 % de la population mondiale utilisait Internet ; alors que dans les pays en développement, la proportion était de 17 % seulement. La fracture numérique demeurait également importante à l'intérieur même des pays développés. Le rapport soulignait de surcroît « l'écart du haut débit » entre ceux qui possédaient une connexion à haute vitesse et ceux qui étaient limités à une connexion via modem⁹.

Lorsque le droit d'accès à l'internet n'est pas garanti ni protégé par l'élimination des barrières structurelles et technologiques, le droit au développement est également menacé, à plus forte raison maintenant que l'accès est perçu comme « un droit fondamental des personnes » par 87 % de ses utilisateurs et par 71 % de ses non-utilisateurs. En d'autres termes, le droit à l'internet peut être considéré comme une composante du droit au développement¹⁰.

Internet et le droit à l'éducation

L'Article 10 de la *Déclaration Universelle des Droits de l'Homme* reconnaît le droit à l'éducation, qui est réaffirmé dans plusieurs autres documents comme dans la *Déclaration Mondiale sur l'Éducation pour Tous* de 1990, dans les Articles 13 et 14 du *Pacte international relatif aux droits*

8 Voir dans cet ouvrage : Adama Samassékou, *Multilinguisme, Objectifs du Millénaire pour le Développement et Cyberspace*.

9 Le rapport précise également : « Le problème, si l'on veut connecter plus de gens dans les pays en développement, c'est que les réseaux à large bande sont en nombre limité. Un grand nombre des applications TIC les plus performantes dans le domaine du développement, comme la télémédecine, le e-commerce, la cyberbanque en ligne et la cyberadministration nécessitent une connexion internet à haut débit. Mais une fracture importante subsiste entre ceux qui ont un accès rapide à un monde en ligne de plus en plus riche en contenu multimédia et ceux qui se débattent encore avec un accès internet commuté et partagé. Fin 2008, la pénétration du large bande fixe était en dessous de 3 % en moyenne dans le monde en développement et se limitait essentiellement à un petit nombre de pays. La Chine – plus gros marché de large bande fixe au monde – compte pour la moitié des 200 millions d'abonnements. Dans la plupart des pays les moins avancés, le nombre d'abonnements au large bande fixe reste négligeable. Les services sont hors de prix et inaccessibles pour la majorité des gens. Cependant, l'introduction de réseaux sans fil à large bande et à haut débit dans un avenir proche devrait accroître le nombre d'utilisateurs internet dans les pays en développement ».

10 Un sondage réalisé pour le compte de BBC World Service en mars 2010 donnait à entendre que 4 personnes sur 5 à travers le monde estimaient l'accès à l'internet comme un droit fondamental <http://news.bbc.co.uk/2/hi/8548190.stm>

économiques, sociaux et culturels et dans l'Article 2 du *Protocole n°1 de la Convention Européenne des Droits de l'Homme*. La communauté internationale accepte également que le droit à l'éducation puisse s'appliquer au-delà du cadre de ces seuls traités.

L'éducation permet non seulement d'acquérir les compétences nécessaires pour encourager la création et soutenir la productivité, mais elle est un facteur de croissance économique. Le savoir n'est plus simplement un dispositif, mais un outil économique dont l'importance ira grandissante dans une économie mondiale de plus en plus déterminée par la connaissance. Cela ne manque pas d'avoir des répercussions sur la distribution des richesses sur les plans mondial et individuel. L'écart grandissant entre l'éducation dans les pays riches et celle des pays pauvres se reproduit au niveau des foyers avec un fossé se creusant entre individus compétents et non compétents¹¹.

Le droit à l'acquisition de compétences dans les technologies de l'information (TIC) est inclus dans le droit à l'éducation, créant ainsi pour les États des obligations positives. Il est également étroitement lié au développement et à l'éradication de la pauvreté. Ces obligations sont exposées dans l'Article 2(2) du *Pacte international relatif aux droits économiques, sociaux et culturels*, incluant la clause relative au droit à l'enseignement primaire accessible à tous établi sur des bases non discriminatoires. La *Convention concernant la lutte contre la discrimination dans le domaine de l'enseignement* (1960) de l'Unesco souligne également ces obligations.

L'internet est l'un des outils les plus efficaces pour s'assurer que l'Objectif 2 des OMD soit atteint. Dans cette perspective d'ici à 2015 de donner à tous les enfants les moyens d'achever un cycle complet d'études primaires, l'enseignement à distance permettrait de rendre accessibles pour un coût minimum des opportunités d'enseignement à un nombre maximal de personnes. L'enseignement à distance évolue et l'internet s'assimile désormais à une salle de classe virtuelle où se déploie une intense interactivité, et un riche partage de ressources et d'informations. Mais pour la plupart des pays en développement, cette technologie reste encore hors de portée à cause d'un manque d'infrastructures et d'éducation à la culture numérique¹². En outre, la plupart des contenus disponibles actuellement en ligne a été conçue en Europe ou en Amérique du Nord et n'est pas

11 A. Clapham, *International Human Rights Lexicon*, Oxford University Press, 2005.

12 <http://www.itu.int/newsarchive/wtd/2001/FeatureEducation.html>

toujours adéquate pour les étudiants ailleurs, tant sur le plan de la langue que des contenus. Cependant le fait que davantage d'universités soient en train de mettre en ligne les programmes existants d'enseignement à distance, démontre son potentiel en tant qu'instrument de déploiement de l'éducation.

Le droit à la diversité linguistique et culturelle sur internet

L'évolution des TIC devrait offrir un potentiel nouveau à toutes les cultures et à toutes les langues. Pourtant la mondialisation du marché et la convergence technologique en cours menacent aujourd'hui le pluralisme culturel et linguistique.

Le premier instrument juridique créé pour la protection des minorités a été la *Convention pour la Prévention et la Répression du Crime de Génocide* (CPRCG) des Nations Unies. Parmi les règles ultérieures protégeant les droits des minorités on doit inclure le *Pacte International Relatif aux Droits Civils et Politiques* (Article 27), la *Déclaration des droits des personnes appartenant à des minorités nationales ou ethniques, religieuses et linguistiques* des Nations Unies¹³, la *Convention-cadre pour la protection des minorités nationales*, la *Charte européenne des langues régionales ou minoritaires* (adoptée par le Conseil de l'Europe), et le *Document de Copenhague* de 1990 par l'Organisation pour la sécurité et la coopération en Europe. De surcroît, plusieurs pays disposent de lois, commissions et institutions afférentes, ainsi que d'institutions pour la protection des minorités, dont parfois des minorités linguistiques.

Avec la *Déclaration universelle sur la diversité culturelle* de l'Unesco (2001) les questions actuelles relatives à la diversité culturelle ou à « l'héritage commun de l'humanité » deviennent des priorités. Cette déclaration fait de la diversité culturelle un impératif éthique concret, inséparable du respect de la dignité humaine.

Si internet est aujourd'hui l'instrument de la mondialisation par excellence, la promotion d'une véritable diversité culturelle et linguistique dans le cyberspace devrait être l'un de ses traits essentiels. Toutes les

13 UN A/RES/47/135, 18 December 1992.

langues et en particulier celle des communautés minoritaires devraient apparaître comme véhicules de culture et de communication en ligne.

Bien que des langues disparaissent, le principe de mettre en avant la diversité culturelle sur internet doit être sauvegardé et ne doit exclure aucune langue. Cette promotion doit encourager le développement de contenus dans les langues sous-représentées, pousser à l'usage de ces langues en tant que langues de travail et aider les communautés en ligne à les utiliser pour communiquer¹⁴. Bref, il faut utiliser tout le potentiel présent et futur de l'internet car il n'y a aucune limite intrinsèque au plurilinguisme.

Droits individuels et accès à Internet : liberté d'opinion et d'expression et liberté d'information

Dans son rapport de 1998 à la Commission des Droits de l'Homme des Nations Unies, le Rapporteur spécial sur la promotion et la protection du droit à la liberté d'opinion et d'expression des Nations Unies s'est montré défavorable à une réglementation d'État de l'accès à Internet et ses contenus :

Les nouvelles technologies, et tout spécialement l'internet, sont profondément démocratiques, assurent l'accès du public et des particuliers aux sources d'information et, à terme, permettront à tous de participer activement au processus de communication. Il souhaite aussi rappeler, qu'à son avis, les mesures prises par les États pour soumettre l'utilisation de ces techniques, et plus particulièrement, une fois encore, l'internet, à une réglementation excessive en faisant valoir que l'accès doit en être limité, réglementé ou refusé afin de préserver la texture morale et l'identité culturelle des sociétés, nient le pouvoir de raison et méconnaissent la capacité de résistance des individus et de la société – à l'échelle d'une nation, d'une province, d'une ville, d'une communauté ou même d'un quartier – d'opérer, dans bien des cas, des ajustements automatiques pour établir l'équilibre sans ingérence ni réglementation excessive de la part de l'État¹⁵.

¹⁴ Marcel Diki-Kidiri, *Towards real linguistic and cultural diversity in cyberspace*. <http://www.portal.unesco.org>

¹⁵ UN Doc. E/CN.4//1998/40IIC4.

Aujourd'hui plus que jamais le pouvoir réside entre les mains de ceux qui sont informés. Les citoyens se révoltent et les régimes sont renversés par le flux d'information circulant à l'intérieur ou entre des pays. Le double rôle d'internet de dissémination et de représentation se révèle de plus en plus important. L'information n'est pas simplement devenue plus puissante, elle est aussi plus accessible¹⁶. Ce qui, historiquement, avait longtemps été le privilège de quelques-uns devient potentiellement désormais l'outil d'une multiplicité.

Le Rapporteur spécial de l'ONU a récemment déclaré à la presse ses inquiétudes face à la législation adoptée en Corée du Sud sur l'internet, en particulier la loi-cadre relative aux Télécommunications et la Loi relative à la promotion de l'information, des réseaux de communication sur internet et de la protection de l'information. La première a justifié l'arrestation d'un blogueur qui a publié dans son blog des critiques envers la politique économique du gouvernement pendant la crise financière; la seconde a été utilisée pour effacer des publications en ligne et punir des personnes qui avaient initié via internet des campagnes appelant les consommateurs au boycott¹⁷.

Pour qu'une société mondiale ouverte puisse voir le jour, les États et les autorités politiques ne doivent pas avoir le contrôle de l'information. Un mécanisme d'auto-régulation se doit d'être établi et la conscience sociale doit naître de processus démocratiques et ouverts.

Le droit à la liberté d'expression est reconnu par l'Article 19 de la *Déclaration Universelle des Droits de l'Homme*:

Tout individu a droit à la liberté d'opinion et d'expression, ce qui implique le droit de ne pas être inquiété pour ses opinions et celui de

16 Selon J. Baudrillard, *Simulacres et Simulation*, Galilée, 1981, « nous vivons dans un monde où il y a de plus en plus d'information et de moins en moins de sens ».

17 UN Doc. À/65/284, par. 76, 11 August 2010. Le rapport signale également que les lois d'urgence ou de sécurité nationale sont souvent utilisées pour justifier des mesures restrictives envers des citoyens journalistiques exprimant leurs opinions ou disséminant l'information via internet, au prétexte souvent de protéger des intérêts nationaux vaguement définis ou de préserver l'ordre public. C'est justement dans ce sens que le Rapporteur spécial des Nations Unies, rejoint en cela par le Président du Groupe de travail sur la détention arbitraire et le Rapporteur spécial sur la torture, le 27 de février 2004, ont envoyé un appel urgent au gouvernement de la Syrie concernant l'arrestation d'une personne due à une série d'articles diffusés par courrier électronique. Les autorités de la Syrie ont répondu que les articles étaient « nuisibles à la sécurité et à la réputation de la nation » et contenaient « plein d'idées et de considérations opposées au système du gouvernement de la Syrie ».

chercher, de recevoir et de répandre, sans considérations de frontières, les informations et les idées par quelque moyen d'expression que ce soit.

Cette disposition inclut le droit de chercher et de transmettre informations et idées, d'informer et d'être informé. Le droit à la liberté d'expression est protégé également par l'Article 19 du *Pacte international relatif aux droits civils et politiques*, par l'Article 10 de la *Convention européenne des Droits de l'Homme*, et relève des normes générales en vigueur du droit international. Là où l'accès au réseau est entravé, ces droits sont violés. La censure étatique est l'une des formes les plus élémentaires de restriction de ces droits; mais la sous-représentation des langues minoritaires constitue de même une barrière à l'avènement de ces droits. En tant que média nouveau et démocratique, internet devrait être un environnement où sont protégés les droits de l'homme.

Au-delà de l'utilisation de l'architecture juridique des Droits de l'Homme pour élargir la participation aux nouveaux médias, internet protège également les Droits de l'Homme dans la mesure où il permet l'accès à l'information. Une meilleure communication peut aussi contribuer à l'application du droit, ainsi qu'à mobiliser pour le changement. D'un autre côté, ces opportunités nouvelles et prometteuses vont de pair avec une responsabilité: Internet n'est pas seulement un outil remarquable de démocratisation, mais peut également servir à répandre la haine et l'abus. Quoi qu'il en soit, la liberté d'expression doit être préservée.

La dernière question est celle de la protection de la vie privée en tant que Droit de l'Homme. Ce droit à la vie privée fut intégré au domaine des droits humains protégés et reconnus dans l'Article 12 de la *Déclaration Universelle des Droits de l'Homme*, interdisant des immixtions arbitraires dans «*la vie privée, la famille, le domicile ou la correspondance*», ainsi que dans l'Article 17 du *Pacte international relatif aux droits civils et politiques*, et dans l'Article 8 de la *Convention européenne des Droits de l'Homme* sous des termes légèrement différents. La *Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel* (1981), émis par le Conseil de l'Europe, connue également comme *Convention 108*, reste, à ce jour, le seul document international applicable à tous les pays, non-membres inclus. Un grand nombre de dossiers traités par la Cour européenne des Droits de l'Homme concernent le droit à la vie privée.

Le stockage de données personnelles relève également de la protection de la vie privée; le Comité des Droits de l'Homme des Nations Unies a précisé que lorsque de l'information est rassemblée et détenue par les autorités publiques ou des entités privées, toute personne doit avoir le droit de déterminer si des informations personnelles la concernant sont incluses, si oui lesquelles et dans quel but elles ont été récupérées (Commentaire Général, Article 17, *Pacte international relatif aux droits civils et politiques*)¹⁸.

LE MULTILINGUISME EN ACTION

Une langue est considérée menacée si elle ne dispose que d'un nombre restreint de locuteurs, d'un statut politique fragile et lorsqu'elle n'est plus transmise à travers l'éducation formelle¹⁹. Des langues disparaissent à un rythme jamais vu auparavant; l'utilisation croissante d'internet pourrait accélérer cette perte des « langues minoritaires » et, par là, de la diversité culturelle/linguistique.

En décembre 2009, les dix langues les plus présentes sur internet étaient : l'anglais (495,8 millions d'utilisateurs), le chinois (407,7 millions d'utilisateurs), l'espagnol (139,8 millions d'utilisateurs), le japonais (96,0 millions d'utilisateurs), le portugais (77,6 millions d'utilisateurs), l'allemand (72,3 millions d'utilisateurs), l'arabe (60,3 millions d'utilisateurs), le français (57,0 millions d'utilisateurs), le russe (45,3 millions d'utilisateurs) et le coréen (37,5 millions d'utilisateurs). Une représentation égale en ligne est essentielle, eu égard au rôle grandissant d'internet en tant que la source première d'information, dépositaire de connaissance, interface de communication et support économique.

Permettre à davantage de gens de communiquer en ligne dans leurs propres langue et écriture est d'une grande importance, en tant que cela permet à diverses populations d'être présentes au sein de l'économie numérique à l'échelle mondiale²⁰. Pourtant, plus de dix ans après l'adoption

18 Voir aussi le jugement de la Cour Européenne sur les Droits de l'Homme dans le cas *Rotaru v. Romania* (Mai 2000).

19 T. Skutnabb-Kangas, *Linguistic Human Rights in Education: Western Hypocrisy in European and Global Language*, article présenté lors de la Cinquième Séance Plénière du Cinquième Congrès International des Études Hongroises.

20 Voir dans cet ouvrage : Michaël Oustinoff, *L'économie des langues*.

d'un modèle architectural multilingue²¹, les langues minoritaires et leurs écritures sont plus que jamais menacés. Le multilinguisme est moins bien établi en ligne que dans « le monde réel »²². Même si l'internet n'est plus un milieu monolingue, parmi les quasi 7000 langues répertoriées de par le monde, très peu ont une présence significative sur internet.²³ Internet a fait ses débuts comme un média principalement monolingue, dans lequel l'anglais était essentiellement la seule langue, et malgré une relative augmentation de la diversité, l'anglais continue d'être la langue dominante de correspondance et de communication²⁴.

Les TIC intègrent de plus en plus des langues et écritures non dominantes, mais si les limitations techniques pouvant empêcher la diversité linguistique en ligne ont été largement résolues, leur présence en ligne est encore un problème²⁵.

La Charte européenne des langues régionales ou minoritaires déclare que,

par « langues régionales ou minoritaires », on entend les langues :
1. pratiquées traditionnellement sur le territoire d'un État par des ressortissants de cet État qui constituent un groupe numériquement inférieur au reste de la population de l'État; et 2. différentes de la (des) langue(s) officielle(s) de cet État

21 Le premier document apportant une architecture mondiale du multilinguisme en ligne date de 1996 pendant l'atelier de travail de IAB (Internet Architecture Board) de l'Institut des Sciences de l'Information (ISI) à Marina del Rey, Californie.

22 C'est intéressant de mentionner les trois catégories de langues dans le cyberspace comme défini par M. Diki-Kidiri: (1) langues de travail utilisées pour communiquer; (2) langues objets que sont simplement citées et étudiées, et (3) langues absentes que ne sont jamais utilisées sur internet.

23 Voir dans cet ouvrage: Daniel Prado, *Présence des langues dans le monde réel et le cyberspace*.

24 P. Lacour, A. Freitas, A. Bénl, F. Eyraud, D. Zambon, *Translation and the New Digital Commons*. <http://lodel.irevues.inist.fr/tralogy/index.php?id=150>

25 WSIS-05/TUNIS/DOC/7-E, 18 novembre de 2005, Article 14, « reconnaît qu'en addition à la construction d'une infrastructure des TICs, il serait convenable d'encourager le développement de la capacité humaine et créer des applications pour les TICs et des contenus numériques dans les langues locales, là où cela serait nécessaire, de façon à assurer une approche compréhensive vers l'objectif de bâtir une Société de l'information mondiale ». Dans l'article 32, les États s'engagent à « promouvoir l'inclusion de tous les peuples dans la Société de l'Information à travers le développement et l'utilisation des langues locales et/ou indigènes dans les TIC » et de « continuer [leurs] efforts pour protéger et promouvoir la diversité culturelle, ainsi que les identités culturelles, à l'intérieur de la Société de l'Information. »

Dans un nombre limité de cas, ces langues bénéficient d'un certain soutien officiel de l'État à travers la loi ou la constitution²⁶. Si langue officielle et langues minoritaires cohabitent dans quelques pays, la cohabitation sur Internet pose un défi plus grand puisqu'un nombre restreint de langues « majeures », et l'anglais en particulier, maintiennent une position dominante, transformant la majorité des langues en minorités.

Toutes les langues sous-représentées ne doivent pas pour autant être considérées comme minoritaires. En premier lieu, aucune langue officielle de l'internet n'a été définie ou reconnue juridiquement. En second lieu, il est nécessaire d'établir des catégories de langues minoritaires ; ce serait une erreur, par exemple, de placer l'araméen, une langue parlée par 2 millions de personnes, au même niveau que le wu, parlé par plus de 77 millions de chinois, et ce malgré une présence équivalente sur internet.

Conséquences d'un nombre limité de caractères sur le web

Le multilinguisme en ligne est important pour : 1. l'efficacité économique et la productivité ; 2. l'établissement d'une société démocratique ; 3. la promotion de l'éducation ; et 4. le respect et la promotion des Droits de l'Homme²⁷. Il constitue donc une manière essentielle de lutter contre la « fracture linguistique » qu'accompagne la « fracture numérique ».

Accepter que des certaines langues soient dominantes sur le réseau a une portée considérable, non seulement pour les autres langues et leurs locuteurs marginalisés, mais pour le développement d'internet lui-même, car une des conditions préalables et essentielles à l'attrait exercé par internet sur ses utilisateurs est celui de pouvoir atteindre une quantité critique de contenus dans une langue donnée. Une représentation linguistique insuffisante en ligne peut avoir un effet dissuasif sur un grand nombre de personnes, même si par ailleurs l'accès lui-même est matériellement possible. Dans la mesure où une fraction considérable de la communication mondiale et des affaires migre sur l'internet, assurer aux milliards

26 Voir dans cet ouvrage : Evgeny Kuzmin, *Politiques linguistiques pour contrer la marginalisation des langues*.

27 Internet Society, *Multilingualism and the Internet (Briefing Paper)*, May 14, 2009. http://www.isoc.org/internet/issues/docs/multilingual-internet-issues_20080408.pdf

d'utilisateurs à venir un accès direct dans leur langue sera d'une importance majeure.

La fracture linguistique est aujourd'hui un frein indéniable pour la majorité de la population mondiale. Des raccourcis peu convaincants, l'utilisation de l'Espéranto ou du Latin comme *lingua franca* par exemple, sont des solutions peu réalistes qui pourraient même rendre la situation plus complexe encore.

Diffuser le multilinguisme et promouvoir la diversité culturelle en ligne, constituent donc un problème à multiples facettes qui requiert la prise en compte d'un grand nombre de paramètres pour atteindre l'objectif de l'internet accessible à tous.

PROMOTION ET PROTECTION DU MULTILINGUISME DANS LE « MONDE RÉEL »

Toute tentative de recherche sérieuse concernant le droit au multilinguisme en ligne requiert de savoir si le multilinguisme est considéré comme un Droit de l'Homme en dehors du « monde virtuel », soit dans le « monde réel ». Langue et Droits de l'Homme apparaissent rarement côte à côte dans les débats, en raison des susceptibilités politiques enchevêtrées aux structures mêmes du pouvoir.

Néanmoins, un examen de la production abondante de textes juridiques de référence depuis la seconde moitié du ^{xx} siècle²⁸, révèle que la plupart des États reconnaissent les droits des minorités linguistiques et acceptent de protéger le droit des personnes à leur langue première, une tendance qui découle en grande partie de l'approche répandue, selon laquelle, garantir le particularisme linguistique assure l'identité culturelle et linguistique de la population.

28 Comme signalé par F. De Varennes dans *To speak or not to speak. The Right of Persons belonging to Linguistic Minorities*, dans un document de travail portant sur le droits des minorités préparé pour le Sous-Comité de l'ONU (21 mars de 1997, <http://www.unesco.org>), des traités internationaux et des conventions portant sur l'usage des langues minoritaires avaient été adoptées depuis le XVI^e siècle. Le Traité d'Union Perpétuelle de 1516 entre le Roi de France et l'État Helvétique contenait ainsi une disposition identifiant ceux qui devaient recevoir certaines protections comme « *les suisses qui ne parlent que l'allemand* ».

Promotion et Protection de la diversité linguistique au niveau international

Alors que les législations nationales ont eu tendance dans le meilleur des cas à tolérer ou à ignorer la diversité culturelle, le droit international et le droit international des Droits de l'Homme ont évolué vers son inclusion. Le droit international des Droits de l'Homme a joué un rôle important en fixant des standards pour les droits linguistiques, constituant ainsi un cadre normatif pour le développement de principes de gouvernance démocratique et de politiques multiculturelles, visant à maîtriser des conflits ethnolinguistique potentiels.

La *Déclaration Universelle des Droits de l'Homme* dans son Article 2(1) déclare que « *Chacun peut se prévaloir de tous les droits et de toutes les libertés proclamés dans la présente Déclaration, sans distinction aucune, notamment [...] (de) langue* ». Le *Pacte international relatif aux droits civils et politiques* (Article 27) propose toujours la protection la plus contraignante pour l'application des droits des minorités linguistiques, déclarant ainsi que « *Dans les États où il existe des minorités ethniques, religieuses ou linguistiques, les personnes appartenant à ces minorités ne peuvent être privées du droit d'avoir, en commun avec les autres membres de leur groupe, [...] ou d'employer leur propre langue* ». L'Article 2(2) du *Pacte International relatif aux droits économiques, sociaux et culturels* place la langue au même rang d'importance que la race, la couleur, le sexe et la religion.

À l'intérieur des cadres du Conseil de l'Europe, de l'Union Européenne et de l'Organisation pour la Sécurité et la Coopération en Europe, les droits des personnes appartenant à des groupes dont les langues sont à caractère régional ou minoritaire ont été abordés dans des traités et conventions multilatéraux.

Aussi bien la *Charte européenne des langues régionales ou minoritaires* (1988) du Conseil de l'Europe que la *Convention-cadre pour la protection des minorités nationales* (1998) possèdent des instances de suivi qui semblent efficaces dans leur mission de pousser les États à s'engager au-delà des exigences minimales.

La *Déclaration des droits des personnes appartenant à des minorités nationales ou ethniques, religieuses et linguistiques*, la *Déclaration universelle des droits linguistiques*, instruments non-contraignants de l'ONU et d'autres

instruments juridiques comme la *Déclaration universelle sur la diversité culturelle* de l'Unesco, sont également pertinents ici, ainsi que le *Projet de recommandation sur la promotion et l'usage du multilinguisme et l'accès universel au cyberspace* (2001) qui a, pour la première fois, défendu la préservation des langues et la diversité au travers de l'accès à des services et ressources électroniques.

Les droits des groupes locuteurs de langues régionales et minoritaires font l'objet de débats au niveau européen depuis 1950. La *Convention de sauvegarde des Droits de l'Homme et des libertés fondamentales* a été suivie par d'autres traités multilatéraux, conventions, accords et un corpus consistant de jurisprudence, comparable à celui codifié par l'ONU, comme la *Charte des droits fondamentaux de l'Union européenne* (2000), la *Charte européenne des langues régionales ou minoritaires* (1992), et la *Convention-cadre pour la protection des minorités nationales* (1994).

La ratification par l'Union Européenne de la *Convention relative à la diversité culturelle* de l'Unesco (2001) apporte une réponse éloquentes aux critiques qui prétendent que l'UE souhaite effacer les caractéristiques nationales ou régionales pour imposer l'uniformité. La diversité linguistique demeure au cœur de l'UE.

Promotion et protection de la diversité culturelle au niveau national

Au niveau national la diversité linguistique acquiert progressivement le statut protégé d'un Droit de l'Homme. Pour une meilleure compréhension des opportunités et des défis auxquels fait face le multilinguisme dans une société multiculturelle, l'Afrique du Sud offre un exemple fascinant. Sa constitution post-Apartheid (1996) est probablement la plus généreuse dans le monde envers le multilinguisme, assurant un statut officiel aux onze langues indigènes existantes dans le pays²⁹. L'Article 6 déclare que :

« 4. *Il incombe au gouvernement national et aux gouvernements provinciaux de réglementer et de contrôler, à travers des dispositions juridiques ou autres, l'utilisation des langues officielles. Sous réserve*

29 Pendant l'Apartheid (1948–1994), l'anglais et l'afrikaans étaient les seules langues à être reconnues officiellement en dépit de la vaste gamme d'autres langues apprises et parlées. Le mythe de l'Afrique du Sud en tant que pays uniquement bilingue, anglais-afrikaans a persisté pendant de longues années.

des dispositions du paragraphe 2, toutes les langues officielles doivent jouir d'une parité de considération et faire l'objet d'un traitement équitable. 5. Le Grand Conseil sud-africain des langues est chargé: (a) de promouvoir et créer des conditions pour le développement et l'usage de: (I) toutes les langues officielles; (II) des langues khoï, nama et san; et (III) de la langue des signes; et (b) de promouvoir et assurer le respect pour (i) toutes les langues communément utilisées.»

D'autres mesures adoptées par certains pays peuvent être mentionnées dans ce qui suit :

En Autriche, au regard de la *Loi relative aux groupes ethniques*, l'administration fédérale est obligée de promouvoir des mesures protégeant l'identité des minorités linguistiques.

Aux Philippines, L'Office des Communautés culturelles du Nord et L'Office des Communautés culturelles du Sud ont été créés pour promouvoir et protéger les droits de personnes appartenant à des minorités linguistiques et à d'autres minorités. Des entités semblables existent dans plusieurs autres pays comme la Russie, le Canada, l'Australie, la Chine et l'Inde.

En Inde la constitution inclut une large reconnaissance des droits des minorités en vue de préserver leur langue, leur écriture et culture (Article 29), le droit d'établir et administrer les institutions d'enseignement dans la langue de leur choix (Article 30), et le droit d'une langue minoritaire d'être reconnue officiellement par les autorités publiques partout où elle représente une proportion considérable de la population (Article 347).

La constitution de la Hongrie contient de nombreuses clauses garantissant les droits des minorités. La Loi du 7 Juillet de 1993, inclut le respect aux Droits de l'Homme pour les membres de groupes minoritaires.

Dans l'Union Européenne des pays voisins comme l'Italie et la France sont deux exemples de politiques opposées ou au moins largement différenciées concernant la protection des minorités linguistiques. L'Italie est nettement davantage engagée dans la sauvegarde du multilinguisme, tandis que la France, plus avancée dans ses positions envers le multilinguisme sur l'internet, pour défendre la « francophonie » ne donne pas un statut légal aux langues minoritaires, lesquelles sont considérées comme langues régionales.

En France, plusieurs dispositions juridiques et réglementaires définissent le rôle de la langue dans la culture, l'éducation et les médias : la *Loi Deixonne* de 1951, la *Loi Haby* de 1975 et la *Loi du 4 août de 1994* (dite *Loi Toubon*). Cette dernière spécifie que « *les dispositions de la présente Loi s'appliquent sans préjudice de la législation et de la réglementation relatives aux langues régionales de France et ne s'opposent pas à leur usage* » (Article 21). La Loi constitutionnelle adoptée le 23 juillet 2008, prévoit que « *les langues régionales appartiennent au patrimoine de la France* » (Article 75-1). La *Charte européenne des langues régionales ou minoritaires* a été signée par le gouvernement français mais n'a pas été ratifiée.

L'Italie, de son côté, a toujours favorisé la préservation de la diversité culturelle sur son territoire. L'italien est la langue officielle mais la législation italienne (lois 482/1999 et 38/2001 ; décret promulgué par le Président de la République 345/2001) et l'Article 6 de la Constitution italienne valorisent les langues non-dominantes³⁰. La loi 482/1999 déclare que ces langues et cultures peuvent être enseignées à l'école, que les documents officiels et actes doivent être bilingues, et qu'une langue locale pourra être utilisée dans l'émission d'informations territoriales. Néanmoins, les langues des immigrants, comme l'arabe et le chinois, ne sont pas prises en compte.

Quelques unes des langues protégées en Italie sont plus communément parlées qu'écrites ; les sites bilingues sont donc rares. Des exceptions concernent les populations des zones frontalières qui parlent l'italien comme deuxième langue, comme celles de la région de Trentin-Haut Adige, germanophones, ou celles de la Vallée d'Aoste, francophones, et dont les sites internet sont souvent bi- ou multilingues, particulièrement ceux d'organismes publics pour lesquels le bilinguisme est une exigence légale. Jusqu'à présent, seules quatre des minorités linguistiques d'Italie disposent de programmes sur la chaîne publique nationale, la RAI : les francophones de la Vallée d'Aoste, les germanophones du Tyrol du Sud, les ladinophones dans les Dolomites et les slovénophones de Trieste³¹.

30 En accord à la loi, les langues et cultures suivantes sont préservées et promues : l'albanais, le catalan, le croate, le français, le provençal, le frioul, l'allemand, le grec, le ladin, l'occitan, le slovène et le sarde (total 2.428,770 locuteurs).

31 Pour une plus ample information, voir Minerva Plus Survey, *Final Plan for using and disseminating knowledge and raise public participation and awareness report on inventories and multilingualism issues: multilingualism and thesaurus*, <http://mek.oszk.hu/minerva/survey/wp3multilingua.doc>

À l'exemple des médias publics, d'autres exemples positifs de mesures protégeant les minorités linguistiques existent. Même si elles ne s'inspirent pas directement du *Pacte international relatif aux droits civils et politiques* (Article 27), elles s'expriment en des termes semblables.

L'article 11(1)(a) de la *Charte européenne des langues régionales ou minoritaires*, se prononce de manière générale quant à la conduite à adopter par l'État envers les médias publics et les langues minoritaires. Les États sont nombreux à reconnaître ouvertement que les besoins des minorités linguistiques ne sont pas satisfaits par l'usage exclusif de la langue officielle, ou celle d'une majorité, dans les médias publics. Dans ces cas, le degré d'usage d'une langue minoritaire dans les médias est un indicatif du poids démographique, des besoins et des intérêts de leurs populations respectives.

MULTILINGUISME EN LIGNE

Pour que le multilinguisme puisse devenir une passerelle permettant de dépasser la fracture numérique, il doit être encouragé en ligne. L'ONU place au tout premier plan la promotion de la diversité linguistique pour atteindre l'objectif d'un accès pour tous à la politique, à l'économie et à la culture³². Le concept central sur lequel repose cette promotion est que toute personne a le droit d'habiter le réseau sans limitations et obstacles. L'internet imaginé comme un bien commun, ou semi-commun, est ainsi entièrement reconnu au niveau international.

Une analyse menée par l'organisation à but non-lucratif Internet Society (ISOC) montre que les progrès relativement lents dans l'établissement d'une diversité linguistique en ligne sont liés à plusieurs types de raisons qui peuvent être d'ordre technique, économique, social et politique.

Raisons techniques. Si les TIC catalysent un accroissement sans commune mesure du libre flux de l'information, elles ont toutefois renforcé des inégalités préexistantes, en particulier, la prédominance de l'écriture en caractères latins. De la fracture numérique issue de ce biais, découlent des incidences sur trois plans : **1** l'impossibilité d'utiliser les TIC en raison du manque de connaissances des codes de caractères linguistiques ; **2** l'impossibilité d'adapter ces technologies à un contexte local ; **3** la

³² Top UN official stresses need for Internet multilingualism to bridge digital divide, 14 décembre 2009, UN News service, <http://www.un.org>

difficulté d'utilisation de ces technologies par des personnes ayant un faible niveau d'instruction. La technologie n'est pas neutre et les TIC sont largement dominées par les sociétés qui les ont créées, conduisant à la prédominance des instruments linguistiques occidentaux/latins. Pour que ces technologies deviennent de culture universelle, elles doivent être pensées de nouveau pour être utilisables par d'autres langues.

Les normes utilisées pour l'implantation du matériel informatique et des logiciels doivent aussi évoluer vers plus d'inclusion. Toutes les nouvelles normes technologiques doivent être réévaluées rigoureusement, pour déterminer leur impact culturel et estimer ainsi leur impartialité envers les groupes sociaux et linguistiques.

Promouvoir l'open source et le logiciel libre pourrait être un autre moyen efficace de surmonter les aspects techniques du problème. Dès lors que des tiers peuvent accéder et modifier directement les codes sources, ils deviennent d'excellents véhicules de localisation ou de régionalisation des programmes³³.

Ces obstacles sont aujourd'hui beaucoup moins importants qu'il y a dix ans. Les difficultés techniques restantes concernent les normes, les outils et les compétences techniques. La plupart des normes au sein d'internet évoluent mais doivent aller encore plus loin. De même, les problèmes de codage qui constituaient l'obstacle le plus important à l'utilisation de nombreux caractères d'écriture il y a à peine dix ans se résolvent à présent avec l'introduction d'Unicode³⁴, mais cette approche est partielle et insuffisante. De surcroît, dans de nombreux pays, la normalisation a été plus lente, particulièrement dans les pays en voie de développement. Par conséquent, certaines normes, comme les configurations du clavier, font encore obstacle au multilinguisme. Quant aux outils, peu de logiciels sont réellement multilingues.

Le développement du contenu reste la plus grande problématique. Si le contenu existant sur internet dans une langue donnée est insuffisant, les locuteurs de cette langue seront peu motivés pour utiliser l'internet. En outre, une grande majorité d'organisations préfèrent publier les informations dans les soi-disant « langues internationales » et développent rarement des contenus dans les langues locales.

33 Voir dans cet ouvrage : Dwayne Bailey, *Localisation des logiciels : open source et multilinguisme numérique*.

34 Voir dans cet ouvrage : Stéphane Bortzmeyer, *Normes de l'internet et multilinguisme*.

LES CONTRAINTES POLITIQUES, ÉCONOMIQUES ET SOCIALES.

Des ressources économiques insuffisantes expliquent également en partie la sous-représentation en ligne des langues des pays en développement. Il n'existe aucune motivation économique pour des développeurs de logiciels à produire des outils logiciels pour des langues parlées dans ces pays, même lorsqu'elles sont parlées par des dizaines de millions de personnes, puisque la demande économique pour ces produits est faible. Les gouvernements locaux et les autres acteurs manquent également de financement pour soutenir des activités visant le développement du multilinguisme sur internet. Des plate-formes de développement d'applications pour téléphones portables sont également extrêmement limitées.

Les responsables internationaux sont depuis longtemps conscients de ce problème. En 2002, l'ancien Secrétaire général de l'ONU lança un défi à la Silicon Valley dans le but de libérer leurs énergies créatives et combler la fracture numérique. L'ONU se trouve dans une position unique pour canaliser l'énergie provenant du secteur privé, des organisations gouvernementales et non-gouvernementales. La définition d'objectifs clairs, de responsabilités, de ressources ainsi que des niveaux de coopération est d'une importance capitale pour faire converger ces différentes cultures organisationnelles.

MULTILINGUISME, LE NOUVEL HORIZON DU CYBERESPACE

Quelles sont les prochaines étapes menant à l'avènement d'un cyberspace multilingue ?

Il est tout d'abord nécessaire de favoriser une approche interdisciplinaire vis-à-vis du multilinguisme, problématique à caractère transversal.

L'une des premiers terrains à bâtir est celui d'une approche commune au regard du droit international, public et privé. Cela suppose que de nouveaux droits importants pour la société de l'information soient identifiés, garantis et protégés.

Les universitaires et les juristes devraient développer des procédés d'arbitrage alternatifs et des mécanismes adaptés de médiation, prenant en

compte la diversité linguistique, à la lumière de l'expérience de la politique universelle de résolution des litiges mise en place par l'Organisation Mondiale de la Propriété Intellectuelle et des autres mécanismes d'arbitrage existants dans le cadre des litiges relatifs aux noms de domaine. Ces principes devraient être fixés afin de renforcer le multilinguisme dans les applications technologiques, telles que les robots d'indexation, les moteurs de recherche, les index de base des données, ou le *Cloud Computing*³⁵.

Si nous supposons évidente la tendance du droit national et international à imposer aux États des obligations positives de protection et de promotion du multilinguisme en ligne comme relevant des Droits de l'Homme, nous affirmons par là même l'obligation de créer en ligne des contenus dans les langues minoritaires.

Cependant, la reconnaissance d'internet comme objet relevant du droit national et international, ayant des droits et d'obligations spécifiques, n'est pas encore évidente. Des questions demeurent, comme celle de savoir à qui incombent les obligations relatives au multilinguisme sur internet et quelles sont les autorités compétentes pour juger et sanctionner les violations à de tels droits.

Il paraît plus que jamais nécessaire de placer la langue au cœur des évolutions de l'internet, ce qui exige avant toute chose l'adoption de principes, d'instruments et de normes juridiques, ainsi que la présence permanente dans l'agenda international de la gouvernance de l'internet la question primordiale de la diversité linguistique.

Note de l'auteur :

Je souhaite exprimer ma plus profonde gratitude à M. Richard Delmas pour m'avoir donné la possibilité de réfléchir sur la thématique complexe associant multilinguisme, Droits de l'Homme et internet. Sans son soutien et sa contribution à la rédaction de l'ébauche de cet article, celui-ci n'aurait sans doute jamais vu le jour. Je suis également très reconnaissante envers Louis Pouzin qui m'a apporté une aide précieuse durant la rédaction de cet article. Enfin, je voudrais adresser un grand « Merci » à mon époux, Erwan Marteil, dont l'infatigable capacité d'analyse a permis de démêler l'écheveau de la problématique et a inspiré un grand nombre des contenus évoqués dans l'article.

35 Pour plus amples détails : <http://www.wipo.int/amc/en> et <http://www.wipo.int/amc/en/domains>

LA GOUVERNANCE DE L'INTERNET À L'HEURE DU MULTILINGUISME

Qui décide du multilinguisme sur internet ? Dans quel bureau autorise-t-on les noms de domaines en chinois ou en arabe ? Qui a voulu que la majorité des pages web soient en anglais ? Cet article vous emmène dans les couloirs de la « gouvernance de l'internet ».

Article original en français.



STEPHANE BORTZMEYER est ingénieur informaticien, plus particulièrement spécialisé en réseaux TCP/IP Il travaille pour l'AFNIC et tient un blog où il est question de technique mais parfois aussi de culture ou de politique mais il s'exprime ici à titre personnel. <http://www.bortzmeier.org>

STÉPHANE BORTZMEYER

LA GOUVERNANCE
DE L'INTERNET
À L'HEURE DU
MULTILINGUISME

Une particularité essentielle de l'internet est son organisation : réseau multi-nations et surtout multi-acteurs, l'internet n'a pas de centre. Il n'existe pas de Président de l'Internet, ni de Conseil des Sages Internetiens qui pourraient prendre des décisions s'imposant à tous. En conséquence, des évolutions qui font l'objet d'un large consensus, comme la nécessaire migration vers le protocole réseau IPv6, ont pris un retard considérable car ces évolutions ne peuvent pas être pilotées de manière descendante. Aucune autorité ne peut dire « *Au 30 avril 2012, tous les réseaux connectés doivent l'être en IPv6* » ou bien « *Au 15 mai 2012, seuls des systèmes d'exploitation sérieux, question sécurité, pourront être connectés à l'internet* ». Les migrations dépendent d'une somme de décisions individuelles et, comme en économie ou comme en écologie, la somme des décisions individuelles est souvent très inefficace, et ne permet pas les changements qui profiteraient à tous, mais nécessitent que chaque acteur fasse un effort, ou engage des dépenses.

QUI DÉCIDE DU MULTILINGUISME SUR L'INTERNET ?

Cette absence de Centre, de Direction Générale, ne veut pas dire qu'internet est une anarchie parfaite, dont le pouvoir serait absent. Il existe au contraire plusieurs centres de pouvoir, dont la légitimité est très variable. En outre, il n'y a pas d'égalité entre ces centres : certains sont bien plus puissants que d'autres et il est évident que Google, Microsoft, le gouvernement des États-Unis, Apple, Baidu, Level 3, Cisco, France Telecom ou Facebook ont davantage de poids que M. Michu, M. Ali, M. Li, une petite agence de création Web installée à Charleville-Mézières, ou le gouvernement du Mali.

Cette situation présente des avantages : elle évite la capture de l'internet par un groupe unique, qui imposerait sa volonté. Certes, un tel mécanisme de décision serait souvent utile (comme dans le cas de la migration vers IPv6 cité plus haut), mais il présenterait également des risques énormes si le Centre ou le Directeur prenait de mauvaises décisions. Si l'internet était dirigé par un groupe proche de l'actuel gouvernement français, le réseau des réseaux serait rapidement stérilisé et réduit à un rôle de distributeur de contenus audio-visuels inoffensifs pour le pouvoir en place. Il faut donc mieux que cette situation ne se présente pas, même si tout acteur de l'internet a au moins une fois dans sa vie regretté qu'une décision qui lui semble indispensable ne puisse pas être prise une bonne fois pour toutes !

Dans le cas du multilinguisme, cette absence de Centre s'est souvent fait sentir : difficulté à assurer le passage de « 85 % de logiciels correctement internationalisés » à « 100 % », par exemple, puisque personne ne peut forcer les programmeurs à faire correctement leur travail, ni les administrateurs système à déployer les bons logiciels.

Il existe beaucoup d'organismes qui gèrent un bout de l'internet, une partie de ce gigantesque système. Il n'est évidemment pas question dans le cadre de ce chapitre de les citer tous. Ils appartiennent à plusieurs catégories :

- Les opérateurs réseau et les fournisseurs d'accès et de services internet, comme Tata ou Comcast. Leur rôle dans les questions de multilinguisme est assez faible ;
- Les producteurs de logiciel, la fondation Mozilla, Apple, les développeurs de Wordpress, etc., ainsi que les gérants des gros services internet très utilisés (moteurs de recherche, plate-formes d'hébergement de blogs, etc.). Leur rôle est cette fois crucial puisqu'un logiciel qui ne permettrait, par exemple, que de produire et publier des textes en alphabet latin, limiterait sérieusement le multilinguisme ;
- Les gouvernements nationaux et les régulateurs officiels variés et divers. Les obligations qu'ils édictent peuvent jouer un rôle dans le développement (ou non) du multilinguisme. Ils peuvent également jouer un rôle d'incitation ou d'encouragement ;

- Les organismes de normalisation et les normes techniques qui sont traités dans le chapitre de ce livre consacré à la normalisation¹;
- Et il ne faut pas oublier bien sûr les utilisateurs. Car, contrairement à bien d'autres systèmes de communication, comme la télévision, l'internet est largement fait par ses utilisateurs. Ce sont eux qui, par exemple, produisent l'essentiel du contenu multilingue sur l'internet, dont le modèle est évidemment l'encyclopédie Wikipédia.

LE RÔLE DE L'ICANN

Lorsque quelqu'un cherche désespérément un Centre à l'Internet, il élit en général l'ICANN (*Internet Corporation for Assigned Names and Numbers*) pour tenir ce rôle. Organisation mondiale, apparemment en charge d'une ressource unique (la racine du système des noms de domaine), elle semble bien située pour cette tâche, ce qui rassure tous ceux que l'absence de Direction Centrale de l'Internet inquiète. Résultat, l'ICANN est souvent l'enjeu de nombreuses luttes politiciennes, alors que des domaines bien plus vitaux de la gouvernance de l'Internet, comme la gestion des adresses IP ne suscitent pas autant d'intérêt.

Pourtant, l'ICANN n'est en rien «le régulateur de l'internet» (formule journalistique courante mais tout à fait fausse). Elle n'est même pas «le régulateur du DNS», comme le montre le chapitre suivant sur les registres de noms de domaine. Tordons un peu le cou aux légendes :

- L'ICANN ne gère pas la racine des noms de domaine. Cette tâche, jugée stratégique, reste la propriété exclusive du gouvernement des États-Unis², via son bras technique, l'entreprise Verisign.
- A fortiori, l'ICANN ne gère pas le reste du DNS (les normes sont établies par l'IETF, les politiques sont décidées par chaque registre) et encore moins les autres activités de l'internet. L'ICANN a un certain rôle (en tout cas en théorie) dans l'allocation des adresses IP.
- Le seul pouvoir réel de l'ICANN est la gestion des TLD états-uniens comme .com ou .org. C'est le seul domaine où le terme de «régulateur» soit approprié.

1 Voir dans cet ouvrage : Stéphane Bortzmeyer, *Normes de l'internet et multilinguisme*.

2 <http://www.ntia.doc.gov/ntiahome/domainname>

En conséquence, le rôle effectif de l'ICANN pour ou contre le multilinguisme est très limité. Le cas le plus souvent cité est celui de l'introduction des IDN (Internartionalized Domain Name) dans la racine (devenu effectif le 5 mai 2010). Mais la décision a été prise par le gouvernement des États-Unis, seul gestionnaire de la racine. L'ICANN a joué uniquement un rôle de retardateur (par de longues études tout à fait inutiles³) et s'est chargé de la communication. La partie technique était réglée depuis longtemps, et plusieurs registres avaient introduit des IDN bien avant la racine.

Aujourd'hui, l'ICANN a encore un rôle de pré-examen des demandes de nouveaux TLD dans la racine, ce qui l'a amené à rejeter les demandes de la Bulgarie⁴ et de la Grèce⁵ au nom du risque de confusion (dans le cas de la Grèce, avec un TLD qui n'existe pas!).

REGISTRES DE NOMS DE DOMAINE

Mais, alors, si l'ICANN ne décide pas du déploiement des IDN, qui le décide? Comme souvent dans l'internet, la décision est largement distribuée. Prenons l'exemple de `.fr` (TLD de la France) : à l'heure où ce texte est écrit, les IDN, qui permettraient des noms de domaines accentués, n'y sont pas encore possibles (mais prévus pour l'été 2012). Qui décide? La réponse n'est pas simple. La France a une loi sur la gestion des noms de domaines mais qui ne spécifie pas la totalité de la politique d'enregistrement. En gros, c'est à l'AFNIC (Association française pour le nommage Internet en coopération) que revient ce rôle. Celle-ci a annoncé le début de la réflexion sur le sujet en 2003⁶, sans résultat immédiat. Lors de la consultation publique sur la gestion de `.fr` en 2008, aucune contribution n'avait été enregistrée sur la question des IDN (alors que plusieurs acteurs de la gouvernance de l'internet en France aimaient critiquer l'ICANN pour

3 Ces études étaient inutiles, car du point de vue technique, un serveur de noms ne fait pas de différence entre un IDN –un nom en Unicode– et un nom traditionnel. Du fait de l'usage de l'algorithme Punycode, les deux sont en ASCII dans la mémoire du serveur de noms. Voir <http://www.icann.org/en/announcements/announcement-28oct07.htm> pour un exemple d'une de ces études, dont le principal intérêt a été de fournir une justification au retard de déploiement des TLD IDN.

4 <http://www.centr.org/main/6079-CTR/version/default/part/AttachmentData/data/Daniel%20Kalchev%20-%20bgidn20110202v3.pdf>

5 <http://svsf40.icann.org/meetings/siliconvalley2011/presentation-swords-confusing-gr-segredakis-15mar11-en.pdf>

6 <http://www.afnic.fr/fr/l-afnic-en-bref/actualites/actualites-generales/2511/show/avertissement-de-l-afnic-sur-les-noms-de-domaine-internationalises-idn.html>

sa lenteur dans le déploiement des IDN). Ce n'est qu'en 2010 qu'un rapport d'un organisme semi-officiel, le FDI (Forum des Droits sur l'Internet), avait plaidé pour une acceptation des IDN dans .fr⁷. Les demandes de domaines accentuée en .fr sont recevables à partir de 2012.

De la même façon, chaque registre de TLD décide de si et quand ouvrir aux IDN. Le TLD du Canada (.ca) n'en a pas. En Europe, les TLD de l'Allemagne (.de, avec une mention spéciale pour le fameux caractère ß⁸), de la Suisse (.ch) et de l'Autriche (.at) ont été dans les premiers (lancement commun dans les trois pays germanophones le 1^{er} mars 2004). Par contre, les TLD des pays utilisant une écriture non latine sont souvent bien plus rapides à enregistrer des IDN, ce qui a été le cas de la Chine continentale (.cn, qui n'avait même pas attendu la sortie officielle de la norme technique). Les pays utilisant une écriture de droite à gauche (comme l'Arabie Saoudite (.sa) ont souvent attendu de pouvoir avoir un nom de domaine (domaine de tête compris) entièrement en Unicode.

INTERNATIONALISATION ET LOCALISATION DES PROGRAMMES

L'utilisateur interagit avec d'autres utilisateurs par le biais de logiciels que, dans l'immense majorité des cas, il n'a pas écrit lui-même. Les capacités ou incapacités de ces programmes ont donc une influence directe sur l'expérience multilingue et il n'est pas exagéré de dire que les fournisseurs de logiciels ont, sur ce sujet, une influence plus directe que la plupart des grands organismes officiels.

Ainsi, début 2011, le système d'exploitation Android⁹ ne gère toujours pas l'écriture arabe (et son sens de droite à gauche). Autre exemple : les programmes, même développés par des français¹⁰, qui ont toujours du mal avec les caractères composés (et qui affichent des messages d'erreur du genre « *unsafe characters* »).

7 <http://www.foruminternet.org/institution/espace-presse/communiqués-de-presse/la-langue-et-internet-le-forum-des-droits-sur-l-internet-publie-une-etude-inedite-2984.html>

8 <http://www.denic.de/en/faqs-about-idns-ss.html>

9 Système développé par Google pour les ordiphones.

10 Lorsque l'auteur de ce texte pointe le matin dans son entreprise, la badgeuse affiche « Stéphane Bortzmeyer », sans accent...

Pourquoi ces faiblesses ? D'abord, il est clair que gérer toutes les écritures du monde représente davantage de travail. Si on refaisait toutes les langues pour faciliter la vie des informaticiens, on se limiterait certainement à l'ASCII. Mais il y a aussi un manque très net de sensibilisation des programmeurs. Ainsi, dans la grande majorité des formations de programmeurs, Unicode n'est pas mentionné ou alors uniquement deux heures vers la fin de l'année dans un cours présenté comme « supplémentaire » (alors que le traitement des chaînes de caractères Unicode devrait faire partie de la formation de base).

Donc, ces programmes ont une grande importance pratique dans la prise en compte du multilinguisme. Mais qui les écrit et qui décide (ou pas) de gérer toutes les écritures possibles ? Il y a un très grand nombre d'entités qui développent des programmes. En simplifiant à l'extrême :

- les grosses sociétés commerciales produisant des produits très répandus, comme Microsoft, Google, Apple, Oracle ou IBM, qui développent ce qui est rentable. La plupart des langues utilisées sur la planète sont très minoritaires et parlées par des personnes ayant un faible pouvoir d'achat ;
- les développeurs de logiciel libre¹¹ qui peuvent être des individus isolés mais également certaines des grosses sociétés citées plus haut (Google et IBM ont une importante activité de développement de logiciel libre). Les motivations peuvent être variées. A priori, la disponibilité du code source et la liberté de l'améliorer, quelles que soient les stratégies commerciales de telle ou telle entreprise et ses priorités, permettent une meilleure prise en compte du multilinguisme dans ces logiciels. Toutefois, ce point est compensé par le fait que les personnes qui utilisent les langues les moins « centrales » n'ont pas plus de compétences techniques que d'argent. Et le développeur de logiciel libre isolé n'a pas forcément les moyens de connaître tous les détails de la gestion du multilinguisme. Un effort d'aide et de sensibilisation dans le domaine des logiciels libres serait donc certainement utile ;
- enfin, il existe également un important lot informel de logiciels développés localement, par un informaticien maison, un stagiaire de passage ou bien une petite société de services. Les auteurs de tels

11 Voir dans cet ouvrage : Dwayne Bailey, *Localisation des logiciels : open source et multilinguisme numérique*.

logiciels sont souvent peu compétents en multilinguisme¹², ils n'ont pas les moyens d'une société comme Microsoft, et le logiciel n'étant pas distribué librement, il ne peut pas être amélioré par d'autres que l'auteur. Ces logiciels sont donc souvent la plaie des promoteurs du multilinguisme.

PRODUIRE DU CONTENU QUI N'EST PAS QU'EN ANGLAIS

La section précédente traitait du rôle des programmeurs et des sociétés de logiciel. Si ceux-ci faisaient parfaitement leur travail, et que tous les logiciels du monde étaient entièrement internationalisés, prêts à gérer n'importe quelle langue et n'importe quelle écriture, tout irait-il bien ? Non, car il faut également que les auteurs de contenu s'emparent de ces logiciels et les utilisent. Si un logiciel de messagerie instantanée permet l'échange de textes en ourdou mais que les correspondants préfèrent s'exprimer en anglais, le multilinguisme en souffrira. De même, si un CMS (Content Management System, Système de Gestion de Contenu) permet d'écrire et de publier des textes dans toutes les langues, cela ne signifie pas que les locuteurs de toutes les langues se saisiront de cette opportunité. De nombreux obstacles se présentent en effet sur leur chemin : contraintes de temps et d'argent, poids de l'habitude, résignation, manque de culture internet et/ou technique, culture socio-politique qui encourage la passivité plutôt que la création, etc.

Ce phénomène de non-utilisation des techniques existantes est d'autant plus marqué que beaucoup d'acteurs institutionnels privilégient le discours à l'action : on organise un colloque sur le multilinguisme, où les orateurs défilent pour déplorer le peu de contenu en langue X ou Y sur le Web, alors que le temps et l'argent auraient pu être mieux dépensés à produire du contenu. Ainsi, pour ne citer qu'un exemple parmi de nombreux autres, la table ronde organisée par le Sénat français le 26 janvier 2011 sur le thème « *La création de contenus culturels dans le monde numérique* » ne comptait pas un seul créateur. La fondation Wikimedia n'avait même pas été invitée. En feuilletant récemment une brochure de défense du français dans une librairie parisienne, où un auteur médiatique critique vigoureusement la domination de l'anglais et le manque

12 C'est sans doute le cas de la badgeuse citée plus haut.

de contenus francophones sur le réseau, j'ai eu la curiosité de chercher sur ledit réseau ce que le personnage en question avait fait pour lutter contre cette tendance : rien. Aucun texte de lui (y compris la brochure en question) n'était disponible sur le Web...

Ne suivons pas le triste exemple de cette table ronde du Sénat et voyons les projets qui sont des succès. Le principal est évidemment Wikipédia. C'est un des plus grands succès du multilinguisme sur internet. En mai 2011, cette encyclopédie collaborative compte 269 éditions actives¹³ incluant des langues comme l'alémanique, l'ouzbek, le kurde et le divehi. Aucun autre site web au monde ne peut annoncer tant de langues disponibles (et ce ne sont pas des traductions, comme on le voit souvent sur les sites commerciaux, mais souvent des contenus originaux). Ainsi, la gestion correcte d'Unicode par le logiciel sous-jacent a bien été utilisée : des locuteurs de nombreuses langues ont saisi l'occasion de *défendre et illustrer* leur langue, en la pratiquant. La barrière est relativement basse puisque contribuer à Wikipédia n'impose pas de compétences techniques pointues. Tout ne se passe pas forcément dans la joie et la bonne humeur (les contributeurs du Wikipédia anglophone se disputent régulièrement pour savoir s'il faut écrire *color* ou *colour*) mais ce n'est pas différent de n'importe quel autre projet humain.

Autre exemple de multilinguisme réussi, le site web de Debian¹⁴, un logiciel libre fondé sur une communauté mondiale dynamique et qui assure la disponibilité du site dans une vingtaine de langues.

LANGUES UTILISÉES DANS LES PROCESSUS DE GOUVERNANCE

Un des aspects problématiques du multilinguisme sur l'internet est le poids de l'anglais. En effet, que ce soit dans les équipes de programmeurs¹⁵ multinationales (qu'il s'agisse de logiciel libre ou bien privé), ou dans les organismes de normalisation ou de gouvernance, l'anglais est la seule langue de travail effective. Parfois, des spectacles comme la

¹³ Une liste est disponible en http://en.wikipedia.org/wiki/List_of_Wikipedias

¹⁴ <http://www.debian.org>

¹⁵ Voir par exemple une intéressante discussion sur la langue à utiliser dans les programmes : <http://stackoverflow.com/questions/1046419/variable-naming-and-team-members-who-speak-another-language>. Comme le note un des participants à la discussion « Something gotta give »...

session d'ouverture d'une réunion sont traduits mais le travail réel se fait en anglais.

Y a-t-il une alternative à ce monolinguisme, qui évite l'incroyable lourdeur et les coûts de, par exemple, l'Union Européenne ou autres organismes peu efficaces? Pour l'instant, malheureusement, il semble que non.

UN PEU D'HISTOIRE

L'histoire de l'Internet peut aider à comprendre la situation de sa gouvernance. Mais attention! Comme toute organisation humaine, Internet a aussi ses légendes et ses jolis récits des origines, qui ne sont pas plus proches de la réalité que les autres récits de création. Donc, au tout début, le réseau qui allait devenir l'Internet, ARPANet, était très centralisé. Il lui manquait une caractéristique essentielle de l'Internet, son caractère multi-organisations. À l'époque d'ARPANet, il était encore possible qu'une autorité décide d'un jour J où toutes les machines devaient avoir adopté telle technique. Le dernier de ces jours J a eu lieu en 1983 lors de l'introduction de la famille de protocoles TCP/IP (IPv4 et TCP dans son état actuel).

L'internet multi-organisations est né à peu près à ce moment. À l'époque, toutes les structures de gouvernance un peu formalisées pesaient sur les épaules d'un seul homme, Jon Postel, qui était contraint de ce fait à jouer un rôle personnel déterminant. La croissance de l'internet, ses changements qualitatifs, ont peu à peu rendu cette situation intenable pour lui-même comme pour l'évolution du réseau. Le gouvernement des États-Unis, qui finançait alors la quasi-totalité de l'internet, s'est lancé dans un bras de fer avec les personnes qui en assuraient la gestion. Ce bras de fer, appuyé idéologiquement sur deux documents, le « Livre Vert », puis le « Livre Blanc », a culminé en janvier 1998 quand Jon Postel a tenté de remettre le contrôle des serveurs racines à l'organisme indépendant IANA (Internet Assigned Numbers Authority). Cette tentative a échoué, car le gouvernement des États-Unis a pris alors le strict contrôle de la racine.

Ce contrôle s'est matérialisé par la création, la même année, de l'ICANN (un organisme qui est donc relativement récent dans l'internet), qui a pris alors le relais de Jon Postel pour assurer les tâches centrales. Mais, comme indiqué plus haut, ces tâches centrales ne représentent qu'une partie de l'internet: même du temps de Postel, ce n'était pas lui qui décidait,

par exemple, si le web devait être déployé ou pas (le web a commencé réellement vers 1991, sans aucune implication d'un organisme central).

Question normalisation, l'IETF, sous sa forme existante, date de 1986. Mais elle n'a jamais eu de réel pouvoir de décision, édictant des normes (les RFC) que les acteurs décident de suivre ou non.

On le voit, la plupart des acteurs et organismes cités sont états-unis. Cela aide à comprendre la prédominance historique de l'anglais et de son écriture, l'alphabet latin. Aujourd'hui, alors que la technique est disponible pour un réel multilinguisme, c'est surtout le rôle dominant des États-Unis dans le monde qui contribue à perpétuer ce poids d'une langue particulière.

GLOSSAIRE

Domain Name System - DNS

Ce terme désigne à la fois le système de noms de domaine, système arborescent permettant de créer des identificateurs comme `cooptel.qc.ca` ou `véliplanchiste.com` et le protocole permettant, à partir d'un tel nom, de récupérer des informations comme l'adresse IP, le nom du serveur de messagerie, etc.

Internationalized Domain Names - IDN

Le terme IDN désigne les noms de domaine exprimés en Unicode comme, par exemple, `سنوات.تونس` pour la Tunisie. On utilise parfois le sigle IDNA (*Internationalized Domain Names in Applications*) pour la technique spécifique employée actuellement, qui passe par une conversion locale en ASCII, avant l'envoi au DNS.

Comme les noms de domaine sont des marqueurs d'identité, très visibles et largement utilisés pour la communication, la possibilité de les exprimer dans sa propre écriture est cruciale. C'est ce qui explique que la question des IDN soit traditionnellement si brûlante.

Internet Corporation for Assigned Names and Numbers - ICANN

Une organisation privée états-unienne qui avait été créée en 1998 par le gouvernement Clinton pour assurer les fonctions auparavant dévolues de manière informelle à Jon Postel. La gestion de la racine a finalement échappé à l'ICANN (elle est gérée par un contrat direct entre le gouvernement des États-Unis et Verisign). L'ICANN gère donc la fonction IANA (les registres autres que la racine), et sert de régulateur à certains TLD notamment `.com` <http://www.icann.org>

Internet Engineering Task Force - IETF

La principale organisation de normalisation sur l'Internet, notamment chargée des couches 3 (routage) à 7 (applications). Elle se signale par sa grande ouverture, ses débats et ses normes (les fameux RFC) étant publics. <http://www.ietf.org>

Top-Level Domain - TLD

Domaine de tête dans un nom de domaine, soit le plus à droite. Dans le nom `google.com`, le TLD est `.com`.

BIBLIOGRAPHIE

MILTON MUELLER. *Ruling the root. Internet governance and the taming of cyberspace*. MIT press. 2002.

PRINCIPES ÉTHIQUES REQUIS POUR UNE PRÉSENCE ÉQUITABLE DES LANGUES DANS LA SOCIÉTÉ DE L'INFORMATION

Il existe de facto une inégalité entre les langues du monde bien qu'elles soient égales en droit et en dignité. Cette situation est le résultat de conditions de vie différentes et souvent inégalitaires. Les 70 langues les plus répandues et les plus parlées dans le monde ne représentent que 1% des langues. Doit-on, cependant laisser pour compte les 360 millions d'hommes et de femmes qui représentent les 6% de la population mondiale qui ne parlent aucune de ces « grandes » langues ?

Article original en français.



MARCEL DIKI-KIDIRI, Centrafricain et Français, est actuellement Consultant en Linguistique Appliquée. Avant sa retraite en 2010, il était Chargé de Recherche au CNRS, dans l'unité mixte Langage, Langues et Cultures d'Afrique Noire (LLACAN : CNRS, INALCO).

MARCEL DIKI-KIDIRI

DE L'INFORMATION
DANS LA SOCIÉTÉ
DES LANGUES
ÉQUITABLE
UNE PRÉSENCE
REQUIS POUR
ÉTHIQUES
PRINCIPES

La vie en société ne peut être harmonieuse et bénéfique pour tout le monde que si elle est organisée de telle sorte que des limites conventionnelles circonscrivent et bornent les libertés individuelles. La codification de ces limites inclut celle des comportements individuels et collectifs en termes de droits et de devoirs, d'interdits et d'obligations, de bien et de mal. La ligne de partage, inscrite dans la conscience de chacun, lui permet de savoir s'il agit en toute honnêteté ou bien s'il est de mauvaise foi, quitte à en assumer les conséquences. L'éthique est, en effet, une question de conscience plus que de loi. Les lois peuvent varier dans l'espace et dans le temps. Il y a même de mauvaises lois qui sont mauvaises parce qu'elles heurtent et révoltent la conscience de tous ceux qui ont encore la capacité de distinguer le bien du mal et le courage de dénoncer le mal, parfois, au péril de leur vie. Là où il n'y a pas de lois basées sur des valeurs éthiques, les plus forts s'emparent de tous les biens à portée de main et n'hésitent pas à ériger en lois leurs mauvaises pratiques. Ils ne sont limités que par d'autres aussi forts qu'eux. La paix est constamment précaire, car chaque fois que quelqu'un se croit suffisamment fort pour imposer sa volonté aux autres, il n'hésite pas à déclencher une guerre. Les plus faibles sont démunis et asservis sans aucun recours possible, si ce n'est l'éveil d'une conscience internationale de toutes les personnes de bonne volonté et assez courageuses pour manifester leur révolte et leur écoeurément devant les situations injustes et insupportables créées par les activités des potentats. L'établissement de lois fondées sur la recherche du bien commun et d'un droit équitable pour chacun est un acte de civilisation qui reconnaît et protège des droits fondamentaux et inaliénables pour les individus comme pour les peuples.

Le respect de ces droits est le premier des devoirs pour tous. Les principes éthiques qui suivent sont requis pour une présence équitable des langues dans la société de l'information :

Toutes les langues du monde sont égales en droit et en dignité, comme ceux qui les parlent

Dans la société de l'information, l'accès à l'information et donc au savoir partagé est un droit fondamental, car l'inégalité dans l'accès à l'information et au savoir engendre une cascade d'inégalités inacceptables dans les conditions de vie des humains. Mais pour que chacun jouisse de ce droit fondamental, toute l'information publique et tout le savoir accumulé par l'humanité doivent pouvoir être rendus disponibles dans toutes les langues du monde afin d'atteindre chaque communauté linguistique. Cela peut paraître utopique, vu que l'on dénombre environ 6800 à 7000 langues dans le monde. Mais cela répond à un premier principe essentiel qui est que *toutes les langues du monde sont égales en droit et en dignité* et peuvent toujours être utilisées pour exprimer toutes les connaissances du monde, du moment que leurs locuteurs le veulent. Si l'on admet l'inégalité des langues comme un fait de nature, alors on n'est pas loin de penser que leurs locuteurs sont intrinsèquement inégaux. Ce qui constitue une discrimination de principe totalement contraire à toute éthique.

Chaque langue est un trésor de l'humanité

Il existe en pratique une inégalité de fait entre les langues du monde bien qu'elles soient égales en droit et en dignité. Mais cette inégalité est le résultat de conditions de vie différentes et elles-mêmes inégalitaires créés, au cours de l'histoire, autant par des relations conflictuelles entre les communautés humaines que par l'isolement de certaines d'entre elles. On sait aujourd'hui que les 70 langues les plus répandues et les plus parlées dans le monde ne représentent que 1 % des langues du monde. Ces langues sont parlées, non seulement comme langues premières, mais aussi comme langues secondes voire troisièmes ou quatrièmes, par 94 % de la population mondiale. Il est donc juste de considérer que si l'information et le savoir étaient disponibles dans ces 70 langues, ils seraient à la portée de la majorité de la population mondiale. Ce serait déjà un énorme progrès vers un multilinguisme mondial à la portée du

plus grand nombre. Doit-on, cependant, laisser pour compte les 420 millions de femmes et d'hommes qui représentent les 6 % de la population mondiale qui ne parlent aucune de ces « grandes » langues ? Bien sûr que non. Il est d'une grande importance pour l'humanité de prendre en compte cette tranche de la population mondiale, car elle parle 99 % des langues du monde. Ces 6 % de locuteurs sont dépositaires de 99 % des langues du monde qui sont autant de systèmes de codage de la pensée, du savoir et de l'aventure humaine. Ce n'est donc pas en raison d'une générosité condescendante envers les minorités qu'il faudrait s'occuper des langues minorées, marginalisées, menacées, en danger ou en voie de disparition. C'est tout d'abord parce que ces langues sont autant de *trésors de l'humanité* dont la connaissance révèle la richesse et la complexité du fonctionnement de l'esprit humain et constitue un système unique de codification du savoir, une vision unique du monde. En protégeant chaque langue pour l'empêcher de disparaître un jour, c'est l'humanité toute entière qui est protégée contre l'ignorance et l'oubli de ce qu'elle est véritablement.

La langue maternelle a une valeur identitaire inaliénable

Le respect porté à la langue maternelle participe du respect porté à ses locuteurs natifs. Ceci est tellement vrai que partout où des peuples ont été dominés, leurs langues ont été systématiquement méprisées, marginalisées et dévalorisées par les dominateurs. « *Ces gens ne parlent pas, ils poussent des cris gutturaux* », au mieux, « *ils n'ont pas de langues, ils ne parlent que des dialectes* »¹. Et partant de ce mépris érigé au rang d'une vérité objective, le dominateur, dans un élan de « civilisation », entreprend d'imposer sa propre langue au dominé. Au Bangladesh, comme à Soweto

1 Il est fondamental de rappeler ici que dans l'antiquité « le terme « *barbare* » est appliqué par les Grecs à tout peuple qui ne parle pas leur langue » <http://fr.wikipedia.org/wiki/Barbare>. En 1912, Carl Meinhof publiait *Die Sprachen Der Hamiten* (Les langues des Hamites). Il utilisait le terme Hamitique. Le système de classification des langues hamitiques était basé sur la croyance que « *les locuteurs des langues hamitiques se sont largement mélangés à des peuples de pasteurs d'origines essentiellement caucasiennes, intrinsèquement différents et supérieurs aux « Nègres d'Afrique »* ». Meinhof a créé une fallacieuse catégorie « Nilo-Hamitique » pour y mettre des langues qui avaient déjà été démontrées et établies comme étant purement nilotiques par Lepsius et de Johnston. Voir http://en.wikipedia.org/wiki/Carl_Meinhof. Voir également Meinhof, Carl. 1906. *Grundzüge einer vergleichenden Grammatik der Bantusprachen*. Berlin : Reimer.

(Afrique du Sud) les peuples dominés ont payé de leur vie le droit de parler leurs langues maternelles! Même dans les cas extrêmes où des peuples déportés, transplantés, déracinés et réduits en esclavage ont été contraints de perdre leurs langues au cours de leur tragique histoire, ils ont su se créer, sur les lieux de leur transplantation, de nouvelles racines, une nouvelle identité sertie dans une nouvelle langue maternelle, une langue créole, construite avec les matériaux de la langue imposée par le dominateur. Le lien qui attache l'être humain à sa langue maternelle est aussi fort que celui qui l'attache à sa patrie, à sa demeure. «*Ma langue est la demeure dans laquelle je vis, la tienne est pour moi comme la fenêtre qui me permet de regarder vers l'extérieur*»². Aussi, est-il clair que nul n'a le droit de priver quelqu'un de sa langue maternelle.

La langue maternelle compte non seulement pour l'individu, mais aussi pour la communauté linguistique qui la parle. La langue est un puissant facteur identitaire pour les communautés humaines quel que soit leur type d'organisation sociale: clans, tribus, ethnies, nations ou autres. La langue est le premier instrument d'intégration de nouveaux arrivants dans une société, qu'il s'agisse d'une nouvelle génération d'enfants ou de migrants. La langue est un facteur de cohésion sociale qui exprime et porte la culture, le savoir, les aspirations, les valeurs de la communauté entière. C'est aussi par l'éducation dans la langue maternelle que toutes ces valeurs se transmettent aux futures générations, permettant ainsi à la communauté de perdurer. C'est parce que la langue maternelle de chaque communauté humaine a une valeur identitaire inaliénable qu'il faut tout mettre en œuvre pour rendre possible à chacun de vivre dans sa langue maternelle et de la faire vivre, y compris dans le cyberspace.

2 Propos attribué à Ogotemmêli s'entretenant avec Marcel Griaule, anthropologue français, auteur de: *Dieu d'eau, entretien avec Ogotemmêli*. (1948) (re-édition Fayard, 1997, 224 pages).

L'accès à l'information et au savoir est un droit fondamental qui doit être garanti à toutes les communautés linguistiques afin d'assurer le partage équitable du savoir

La masse des informations et des savoirs du monde est immense et, dans le cyberspace, les informations qui s'y trouvent ont une croissance exponentielle. Il serait donc illusoire de chercher à rendre dans les 7000 langues du monde chaque information ou élément de connaissance publié dans le cyberspace. Ce n'est donc pas ce qui est préconisé ici. La vision mise en avant est bien plus pragmatique et réaliste. Dans le cyberspace, tous les peuples du monde sont « voisins », car ni le temps ni la distance géographique ne constituent une frontière. Tout ce qui est en partage gratuit est à la portée de toute personne susceptible d'y trouver un intérêt. La mise en commun gratuite de toutes les connaissances, comme en témoigne une encyclopédie comme Wikipédia et bien d'autres bases de données en ligne, a été pendant longtemps, et demeure encore largement, ce qui caractérise et distingue le réseau mondial. Au fur et à mesure que le commerce électronique se développe sur internet, la question des droits d'auteurs pour les créateurs (musiciens, cinéastes, designers, écrivains, etc.) devient cruciale et s'accommode mal du téléchargement gratuit des produits culturels et intellectuels. D'où le besoin d'une réglementation faisant la part du payant et du gratuit afin de protéger à la fois les droits légitimes des auteurs, artistes et créateurs, d'une part, et, d'autre part, la liberté des échanges qui caractérise le cyberspace. Mais si on n'y prend garde, le besoin de réglementation peut facilement dériver vers une réduction croissante des espaces de liberté et ainsi priver les plus démunis des moyens de participer à la vie de la Société de l'information. Une chose est certaine : il n'est plus nécessaire de comprendre l'anglais pour utiliser un ordinateur, de même qu'il n'est plus nécessaire d'avoir un diplôme en informatique. L'anglais n'est plus la langue obligée de l'utilisateur de l'ordinateur et il n'y a plus vraiment de langue indispensable, mais seulement les langues personnelles des utilisateurs. Alors, parce que l'accès à l'information et au savoir est un droit fondamental qui doit être garanti à toutes les communautés linguistiques afin d'assurer le partage équitable du savoir, il est très important que, dans le respect de la propriété intellectuelle, chaque communauté linguistique ait le droit de puiser dans ce savoir tout ce dont elle a besoin et qu'elle estime nécessaire de rendre dans sa langue.

Les solutions techniques qui facilitent le multilinguisme doivent être privilégiées à tous les niveaux d'intervention sur les moyens de communication impliquant les langues

La société de l'information est caractérisée par un usage massif et mondial des technologies de l'information et de la communication. La structure de ces technologies doit donc évoluer de façon à rendre possible l'utilisation de n'importe quelle langue comme support de communication, à l'oral comme à l'écrit. Aussi, l'établissement de normes standards internationales est d'une grande importance pour faciliter l'utilisation des langues avec les TIC³. Le système de codage Unicode en est un des meilleurs exemples. Si l'ASCII étendu à 8 bits peut coder un maximum de 256 caractères, Unicode, lui, permet de coder actuellement plus de 110 000 caractères graphiques. Il est donc, en principe, capable de traiter informatiquement tous les systèmes d'écriture du monde. Un autre exemple significatif est celui du codage des langues (ISO 639-*n*) où le *n* est un chiffre allant de 1 à 6 et représentant les différentes versions de la norme. Ce système de codage des noms de langues et groupes de langues est fortement concurrencé par le système de référencement des langues du monde, LS 640, mis au point par David Dalby de l'Observatoire de la *Linguasphère*, et compatible avec la norme ISO 11179. Enfin la norme ISO 15924 permet de référencer tous les systèmes d'écriture du monde avec leurs différentes variantes diachroniques. Toutes ces normes à visée mondiale constituent autant de fondations pour la prise en compte de toutes les langues du monde dans le processus de développement des technologies de l'information et de la communication. Les problèmes techniques sont, certes, toujours complexes, mais le principe éthique qui devrait prévaloir ici est qu'à tous les niveaux d'intervention, on doit privilégier les solutions techniques qui assurent une plus grande flexibilité et une plus grande ouverture sur des applications multilingues. Comme le dit Jean-Louis Garçon⁴:

3 Voir dans cet ouvrage l'article de Stéphane Bortzmeyer, *Normes de l'internet et multilinguisme*.

4 GARÇON Jean-Louis 2005 *Ntic & Éthiques... Quelle valeur possède l'information en ligne?* Mémoire de DESS en Ingénierie multilingue, Centre de Recherche en Ingénierie Multilingue, INALCO.

Les instruments sont maintenant plus ou moins en place. Ils ne sont pas encore parfaits, mais on peut désormais surfer sur le Web en utilisant le chinois, le japonais, le coréen, et de nombreuses autres langues qui n'utilisent pas l'alphabet occidental. Comme Internet s'étend à des parties du monde où l'anglais est très peu utilisé, par exemple la Chine, il est naturel que ce soit le chinois et non l'anglais qui soit utilisé. La majorité des usagers en Chine n'a pas d'autre choix que sa langue maternelle (p. 85).

Le développement d'infrastructures de communication dans les communautés les plus défavorisées est un devoir de solidarité et d'équité

Parmi les questions fondamentales, l'insuffisance ou l'absence de l'équipement en infrastructures de communication est un facteur important dans l'impossibilité d'utiliser les TIC par les communautés linguistiques défavorisées. Dans les pays industrialisés, l'écart entre la ville et la campagne est généralement très important en matière d'infrastructures de communication. Cet écart est encore plus dramatique dans les pays en développement car il se conjugue ici avec d'autres écarts tout aussi graves en matière de services publics.

Le déploiement d'infrastructures lourdes (fibre optique notamment) dans les pays en développement tout comme dans les zones rurales des pays développés est un devoir de solidarité qui permet aux communautés les plus défavorisées d'accéder à la communication et donc à l'information et au savoir partagé. Grâce à cette solidarité, des communautés linguistiques peuvent développer les outils nécessaires à l'usage de leur(s) langue(s) dans l'espace de communication de la société de l'information. Le devoir de solidarité envers les communautés défavorisées est ainsi l'un des principes éthiques nécessaires à assurer la présence de langues maternelles dans la société de l'information. Il s'agit ici aussi bien d'une solidarité nationale qu'internationale.

Sur le plan national, il revient au gouvernement et au législateur de garantir que la solidarité nationale puisse jouer en choisissant de développer non seulement les agglomérations urbaines mais aussi les campagnes en matière d'équipement en infrastructures de communication. Dans

un pays multilingue, cela revêt une importance accrue car c'est une des conditions de base pour que les langues régionales puissent accéder localement au cyberspace.

Sur le plan international, le fait même pour des pays industrialisés, des organisations internationales ou des entrepreneurs privés étrangers d'investir ou de se faire l'avocat des investissements dans les pays en développement afin de les aider à réaliser leurs plans nationaux d'équipement en infrastructures de communication participe du principe de solidarité si les conditions de ces investissements prennent en compte les intérêts véritables des pays en développement. Il faut souligner très clairement que dans tous les cas l'investissement nécessaire est important et donc ne peut pas s'envisager comme une œuvre humanitaire sans contrepartie économique. Bien au contraire, c'est toujours une opération hautement économique et donc rentable pour les investisseurs. Et c'est bien pour cela qu'il faut rappeler fortement le principe éthique de solidarité et d'équité afin que la logique de la rentabilité économique ne soit pas la seule à prévaloir dans les décisions politiques relatives au développement de la communication. Il doit être clair aussi que l'on n'étend pas un réseau de communication sur tout un territoire national ou tout un continent sans se préoccuper de la formation des hommes, aussi bien des hommes de métier (techniciens, gérants de réseaux, etc.) que des consommateurs qui auront à utiliser ces nouveaux modes de communication.

La capacitation des communautés locales par la formation des hommes

L'organisation de la capacitation des communautés linguistiques nécessite également d'importants investissements et un programme de longue durée. Il faut former davantage de personnels à mesure que le réseau s'étend ou est envisagé sur tout le territoire national. Il faut aussi faire évoluer les habitudes de communication des utilisateurs du réseau en facilitant l'accès aux terminaux que sont les ordinateurs, les téléviseurs, les téléphones mobiles, etc. Sans ce marché, l'utilisation des réseaux ne serait pas suffisamment intense pour les rentabiliser. La capacitation des spécialistes comme des usagers est donc à la fois une nécessité économique et un droit humain autant qu'une condition de réussite. C'est pourquoi il est important qu'elle soit fortement soutenue, sinon entièrement prise

en charge, par une action gouvernementale vigoureuse. Les associations professionnelles ont intérêt à développer les cybercafés qui sont autant de lieux propices à l'accès à l'internet quand on n'a pas d'ordinateur chez soi. Plus les populations ont la possibilité d'accéder aux terminaux de communication et sont capables de s'en servir, plus se développent les conditions propices à l'utilisation de leurs langues locales dans les TIC. Il faut donc constamment solliciter et rechercher l'engagement des gouvernements pour une politique linguistique favorisant le multilinguisme afin que chaque langue maternelle parlée sur leurs territoires soit valorisée, et utilisée dans les réseaux de communication.

Les autorités gouvernementales ont un rôle irremplaçable à jouer dans la promotion du multilinguisme impliquant les langues des populations locales et les TIC

Le 11 janvier 1993, Al Gore, le vice-président des États-Unis, lançait devant la presse le projet global des « autoroutes de l'information ». Internet s'est d'abord développé en Amérique du Nord et en Europe pour atteindre ensuite l'ensemble des pays industrialisés. Ce n'est que dans un second temps que le réseau internet se propage lentement dans les pays du Sud ou les pays en développement. Sans l'implication importante des gouvernements, le réseau des réseaux n'aurait sans doute jamais vu le jour. L'installation et le déploiement d'un réseau de communication et d'information sur tout un pays, voire tout un continent, ouvre des opportunités infinies d'intensifier des activités économiques et culturelles et d'accélérer l'évolution sociale. Cela aiguise forcément les appétits. Les gros investisseurs tant privés que publics, généralement des pays industrialisés, ne manquent pas de tenter d'évincer les gouvernements des pays en développement du contrôle des investissements qu'ils acceptent de faire dans ces pays pour l'installation des réseaux de communication. Là où ils y parviennent, ils dictent les prix sans tenir compte du pouvoir d'achat de la population locale mais dans le seul souci de rentabiliser rapidement leurs investissements. Cela crée des inégalités sociales supplémentaires entre ceux qui ont les moyens d'accéder aux TIC et ceux qui ne les ont pas. Ces monopoles soulèvent évidemment des problèmes éthiques graves.

Seuls les gouvernements peuvent faire des choix politiques de développement des TIC intégré dans les plans nationaux de développement économique, culturel, éducatif et social, car seuls les gouvernements ont la responsabilité ultime du destin de leurs pays. Seuls les gouvernements peuvent intervenir pour réguler durablement les prix de façon à garantir le pouvoir d'achat des consommateurs. Seuls des choix politiques gouvernementaux peuvent orienter efficacement les actions économiques, éducatives, culturelles, sociales et communicatives dans un sens qui soit favorable à l'épanouissement de toutes les communautés linguistiques de leurs pays en promouvant un multilinguisme convivial. En effet, une politique gouvernementale bien conçue qui fait une place équitable à chaque langue parlée dans un pays est une politique qui garantit une paix linguistique durable, contribuant ainsi à une paix sociale permanente.

Il n'en demeure pas moins que l'ensemble des forces vives qui constituent autant de centres de pouvoir, à savoir, les organisations nationales ou internationales, les organisations gouvernementales internationales, les associations qui s'impliquent dans des actions culturelles et linguistiques, ont le devoir moral d'assumer leur part d'engagement en se mobilisant partout où cela est nécessaire pour promouvoir le multilinguisme sans exclusive, le développement des technologies qui le portent, et la capacitation des locuteurs.

Les organisations nationales et internationales, gouvernementales et non gouvernementales, ont un rôle important à jouer dans la promotion du multilinguisme dans la société de l'information.

Les organisations internationales ont toujours joué un rôle primordial et d'avant-garde dans la lutte contre les fléaux qui frappent l'humanité. Après les guerres, aussi bien la Société des Nations que l'Organisation des Nations-Unies ont eu comme mission fondamentale la prévention des conflits en s'attaquant aux différentes causes profondes susceptibles de générer des conflits entre les États, notamment, la pauvreté, la faim, l'illettrisme, les grandes endémies, les inégalités économiques, les crimes contre l'humanité, les génocides, etc. Le Programme des Nations-Unies pour le Développement (PNUD) soutient les États dans la lutte contre la pauvreté; l'Organisation des Nations Unies pour l'Alimentation et

l'Agriculture (FAO) lutte contre la faim dans le monde; l'Organisation des Nations Unies pour l'Éducation, la Science et la Culture (Unesco) œuvre contre l'illettrisme, l'analphabétisme, et pour le développement de la science, de l'éducation et des cultures dans le monde. Dans leurs domaines spécifiques, l'Organisation Mondiale de la Santé (OMS), l'Organisation Mondiale du Commerce (OMC) tentent respectivement de juguler les grandes endémies d'une part et de réguler les échanges commerciaux d'autre part. Et si les plus grands criminels ne peuvent pas être poursuivis dans leurs pays, la Cour Pénale Internationale est le dernier recours pour les victimes en quête de justice. Tout cela montre à quel point les organisations multilatérales sont importantes dans pratiquement tous les domaines afin d'agir dans le sens des valeurs éthiques universelles pour lutter contre divers fléaux de la société que l'on peut globalement considérer comme le mal.

C'est donc tout naturellement que nous considérons les organisations internationales qui ont pour mission de promouvoir les langues et les cultures du monde comme des partenaires essentiels dans la protection de la diversité linguistique tant dans la logosphère que dans le cyberspace. Tout comme la biodiversité est indispensable au développement de la vie sur la terre, la diversité linguistique est essentielle à la vie des êtres humains, car c'est à travers sa langue que chaque communauté humaine exprime sa culture, son identité et s'adapte à son milieu de vie et apprend à survivre dans cet environnement. Aussi, à l'heure de la mondialisation où les technologies de l'information et de la communication se répandent dans toutes les communautés humaines quels que soient leurs cadres de vie, de nombreuses cultures fragilisées sont sérieusement en danger de disparition ainsi que les langues qui les expriment et, bien entendu, l'identité des gens qui les parlent. C'est la question existentielle que se posent certains habitants de l'archipel des Tuvalu, conscients et soucieux de la menace de submersion que fait peser sur leur pays le réchauffement climatique. Si les cent mille habitants de l'archipel doivent être évacués vers divers pays comme la Nouvelle-Zélande, l'Australie, les États-Unis, le Canada, la France, l'Angleterre, etc., où ils constitueront de petites minorités appelées à se dissoudre en quelques générations dans leurs diverses sociétés d'accueil, que deviendront l'identité nationale / identité collective des Tuvaliens, leur culture ancestrale, leur langue, leur savoir-faire, etc.? Cette question, on peut la poser également pour les locuteurs des 2400 langues en danger du monde recensées dans l'*Atlas des langues*

en danger de l'Unesco. Les organisations internationales sont le dernier rempart contre l'indifférence dans laquelle peuvent mourir 99 % des langues du monde parlées par seulement 6 % de la population mondiale.

Dans un rapport de l'Unesco du début des années 50, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur internet dans sa langue maternelle pourrait bien être son équivalent à l'Âge de l'Information. Si internet doit vraiment devenir le réseau mondial qu'on nous promet, tous les usagers devraient y avoir accès sans problème de langue. Le considérer comme la chasse gardée de ceux qui, par accident historique, nécessité pratique ou privilège politique, connaissent l'anglais, est injuste à l'égard de ceux qui ne connaissent pas cette langue.

Lancé en janvier 1999 par la Commission européenne, le site HLTCentral (HLT: Human Languages Technologies – Technologies des langues humaines) propose une courte définition de l'ingénierie du langage⁵:

L'ingénierie du langage permet de vivre en toute convivialité avec la technologie. Nous pouvons utiliser notre connaissance du langage pour développer des systèmes capables de reconnaître à la fois la parole et l'écrit, de comprendre un texte suffisamment en profondeur pour être capable de sélectionner des informations, de le traduire dans différentes langues et de générer aussi bien un discours oral qu'un texte imprimé. L'application de ces technologies nous permet de repousser les limites actuelles de notre utilisation du langage. Les systèmes à commande vocale sont appelés à jouer un rôle prépondérant et à faire partie intégrante de notre vie quotidienne.

Puisque des solutions techniques existent⁶, il est possible de les appliquer à toutes les langues du monde si les principes éthiques énoncés et commentés ici sont appliqués par tous les acteurs sociaux, chacun dans sa sphère d'activité.

5 Cette information ainsi que la citation qui suit sont tirées de Marie Lebert (1999, actualisée en 2009) « L'internet et les langues » rapporté par *Le Net des études françaises*: <http://www.etudes-francaises.net/entretiens/multi.htm>

6 Voir dans cet ouvrage: Joseph Mariani, *Les technologies de la langue en soutien au multilinguisme*.

L'INTERNET EN CHINE

La faille qui sépare l'internet chinois de l'internet américain, ou plus généralement occidental, est l'une de celles qui connaît des secousses régulières, parfois même des séismes dont l'impact tant médiatique que politique suscite des réactions parmi les opinions publiques comme au plus haut niveau des États. Entre divergences et convergences, deux mondes se rencontrent sur et par la Toile, ce qui va grandement influencer sur l'organisation future d'internet.

Article original en français.



STÉPHANE GRUMBACH, directeur de recherche à l'Inria, est un spécialiste des bases de données. Il a assumé diverses responsabilités dans la coopération internationale, comme directeur des relations internationales de l'Inria, conseiller scientifique à l'ambassade de France en Chine, et directeur d'un laboratoire conjoint au sein de l'Académie des sciences de Chine, où il a été le premier étranger habilité à diriger des doctorants.

STÉPHANE GRUMBACH

L'INTERNET EN CHINE

L'internet rapproche des individus arbitrairement éloignés sur la planète comme aucun autre média de par le passé. Il leur permet aussi bien de communiquer directement, que d'accéder aux informations qu'ils mettent en ligne, en transcendant parfois même les différences linguistiques grâce aux outils de traduction en ligne. Pour autant, internet est loin d'être un espace uniforme dans lequel l'information circulerait sans contrainte. C'est au contraire un espace fracturé, dont les failles révèlent à la fois la diversité, la pluralité des usages et plus généralement des cultures, mais également les rivalités des différents acteurs pour le contrôler, pour maîtriser les contenus, et dominer les marchés.

La faille qui sépare l'internet chinois de l'internet américain, ou plus généralement occidental, est l'une de celles qui connaît des orages réguliers, parfois même des tempêtes dont l'impact tant médiatique que politique est important, suscitant des réactions à la fois dans les opinions publiques et au plus haut niveau des États. L'équilibre des forces en présence devrait contribuer à orienter l'organisation future de la Toile, ses options culturelles, technologiques et économiques, et le degré d'ouverture ou de cloisonnement, voire de balkanisation du réseau mondial.

Si des différences essentielles se sont fait jour sur la philosophie du réseau, en particulier autour des questions de neutralité du réseau et de liberté d'expression, qui font l'objet à l'Ouest d'une couverture médiatique extrêmement importante, il est clair néanmoins que ce sont les aspects économiques et stratégiques, beaucoup moins médiatisés, qui dominent les mouvements de divergence et de convergence des deux principales plaques de l'internet.

En croissance continue depuis la politique d'ouverture mise en place par Deng Xiaoping en 1978, deuxième économie depuis qu'elle a dépassé le

Japon en 2010, la Chine détient aujourd'hui le tiers des réserves de change mondiales. Elle voit rentrer dans ses universités autant d'étudiants que les États-Unis et l'Europe réunis, et ses chercheurs publient aujourd'hui, dans certains domaines de l'ingénierie, plus que les étatsuniens. Aussi, l'émergence de la Chine comme l'une des premières puissances de la planète assure à ce pays une capacité d'action internationale, en particulier dans les domaines innovants, très forte et trop longtemps sous-estimée.

Dans le domaine de l'internet, la Chine est loin d'être à la traîne. Avec plus de 600 millions d'utilisateurs de mobiles, dont près de la moitié accède au réseau internet via leur mobile; avec 420 millions d'internautes en juin 2010 selon le CNNIC¹, dont 60 % ont moins de trente ans et une appétence très forte pour les nouvelles technologies, la Chine constitue le plus grand marché mondial pour les industries des technologies de l'information. Les multinationales de ces domaines ne s'y sont pas trompées en installant massivement aujourd'hui leurs nouveaux laboratoires de recherche en Chine.

C'est également dans le domaine culturel que le poids de la Chine est immense. Le chinois, ou encore l'ensemble des langues chinoises, est en effet le deuxième groupe linguistique en ligne avec 450 millions d'internautes, juste derrière l'anglais qui compte 536 millions², mais avec un taux de pénétration de seulement 32 %, inférieur donc à celui de l'anglais qui est de 42 %, ce qui lui promet un potentiel de croissance supérieur. Le chinois est donc en passe de devenir une des langues prédominantes de la Toile.

La question de la censure a conduit en 2010 à un bras de fer sérieux entre la Chine et les États-Unis, avec comme principal enjeu la société Google. Suite à des cyber-attaques venant de Chine dont elle s'est dite victime en janvier 2010, Google a décidé d'interrompre la censure qu'elle imposait jusqu'alors à son moteur de recherche sur le territoire chinois en redirigeant google.cn vers google.com.hk, son site à Hong Kong, qui offre une recherche en chinois non censurée. On mesure aisément qu'au-delà des questions de liberté d'expression, les questions commerciales sont loin d'être marginales dans ces épreuves de force.

1 <http://www.cnnic.net.cn>

2 <http://www.internetworldstats.com>, June 2010.

Voir également dans cet ouvrage : Daniel Prado, *Présence des langues dans le monde réel et le cyberspace*.

Si les européens se sont habitués à recourir massivement à des services en ligne offerts par des sociétés américaines (Google, Facebook, eBay, Twitter, etc.), les chinois ont développé des champions nationaux de la toile qui détiennent dans leur pays les plus grandes parts de marché, et occupent quatre des vingt premières positions du classement mondial d'Alexa³, rivalisant ainsi avec les champions américains ! C'est le cas en particulier du moteur de recherche Baidu qui possède plus de 60 % du marché en Chine, et arrive en 6^e position du classement. Baidu offre de nombreux services dont une encyclopédie construite de manière coopérative à la manière de Wikipédia, et se développe également au Japon, mais ironie du sort, le site baidu.jp est inaccessible de Chine.

Le dynamisme des forums, des blogs, des réseaux sociaux, des outils de messagerie instantanée ou de commerce en ligne témoigne à la fois de l'engouement de la jeunesse chinoise pour le réseau, ainsi que de l'efficacité du secteur industriel national. Parmi les sociétés les plus en vue, il convient de citer Tencent, avec son célèbre service de chat QQ, ses jeux, ses mondes virtuels (toute une culture en Chine !), qui est en 8^e position ; Alibaba et ses filiales pour le commerce électronique, dont Taobao, 13^e position ; ainsi que sina.com, pour les informations sur la Chine, en 16^e position.

La suspicion entre les deux régions ne s'arrête pas aux contenus ou aux questions d'ouverture du marché aux sociétés étrangères, mais concerne également la manière dont les données transitent sur le réseau. Au printemps 2010, le fait qu'une partie du trafic provenant de l'administration américaine ainsi que de compagnies privées ait transité pour quelques minutes par la Chine suite à des erreurs de routage, provoquées par l'émission d'informations de routage incorrecte par un prestataire de service internet chinois, a suscité une interrogation forte aux États-Unis sur les causes d'un tel détournement.

L'exemple de la Chine nous montre l'articulation entre politique de développement, présence sur l'internet dans la langue et avec l'écriture nationale, enjeux économiques et géopolitiques. La mondialisation du réseau est devenue réalité, et avec elle la présence dans l'univers numérique des enjeux traditionnels des relations internationales et la définition d'une nouvelle articulation des cultures et des modes de vie.

³ <http://www.alexa.com/topsites>

L'ÉCONOMIE DES LANGUES

Il faut attendre la fin du xx^e siècle pour voir les économistes s'intéresser sérieusement aux langues en tant que domaine d'étude propre pour leur discipline. Ce n'est qu'avec la mondialisation et la montée en puissance du multilinguisme que les langues sont véritablement apparues pour ce qu'elles sont : un enjeu stratégique majeur dans la communication mondialisée, non seulement sur le plan politique, culturel ou sociétal, mais également économique. Un produit n'est plus commercialisable à l'échelle planétaire sans prendre en compte cette dimension : « *No translation, no product* ».

Article original en français.



MICHAËL OUSTINOFF, maître de conférences habilité à diriger des recherches en traductologie à l'Institut du monde anglophone de l'Université Paris 3 Sorbonne Nouvelle, est actuellement en délégation à l'ISCC, l'Institut des sciences de la communication du CNRS. Son troisième ouvrage, *Traduire et communiquer à l'heure de la mondialisation* est paru chez CNRS Éditions en 2011.

MICHAËL OUSTINOFF

DES LANGUES L'ÉCONOMIE

I faut attendre la fin du xx^e siècle pour voir les économistes s'intéresser sérieusement aux langues en tant que domaine d'étude propre pour leur discipline. Pourquoi si tard ? Plusieurs raisons peuvent être avancées, mais on commencera par celle qui est, sans conteste, à la fois la plus évidente et la plus déterminante : parce qu'il est désormais impossible de faire l'impasse sur les langues à l'heure de la mondialisation. Du point de vue économique, les langues sont devenues un secteur industriel à part entière considérable, voire stratégique.

Dans une étude récente sur la taille de l'« industrie des langues » [RINSCHKE 2009], la Direction générale de la traduction (DGT) de la Commission européenne estime sa valeur en 2008 à 8,4 milliards d'euros, constatant que c'est un des secteurs qui ont le mieux résisté à la crise, avec un taux de croissance impressionnant de 10 % (qui dit mieux ?) censé l'amener à peser 16,5 milliards d'euros (estimation basse) en 2015, et sans doute même 20 milliards (estimation haute). Quand on sait que l'évolution est comparable à l'échelle de la planète entière [DWYER 2010], on comprend que les économistes soient amenés à s'y intéresser de plus en plus près, et ce à tous les niveaux de l'analyse économique. Une question fondamentale que les États-Unis prennent désormais au sérieux, tout comme le Royaume-Uni à travers des institutions telles que le British Council [GRADDOL 1997] ou la British Academy [BRITISH ACADEMY 2009].

Néanmoins, l'apparition d'un nouveau secteur industriel dont l'expansion spectaculaire est liée à celle des TIC (Technologies de l'Information et de la Communication) dans le cadre de la mondialisation ne se limite pas à l'analyse de son poids économique, pour incontournable que soit cette question : l'économie des langues est inséparable de sa dimension politique à l'échelle de la gouvernance mondiale.

Replacées dans cette perspective, les langues ne sauraient plus être considérées comme un simple coût – qu’il s’agissait jusqu’ici de réduire au minimum, fonction que le recours au tout-à-l’anglais était censé remplir¹ – mais comme un investissement, aussi bien d’ordre économique que culturel et géopolitique, dont l’intérêt apparaît d’autant plus crûment sur fond de crise financière mondiale généralisée, la plus grave que l’on ait connue depuis l’époque de la Grande Dépression, née, elle aussi, aux États-Unis.

NO TRANSLATION, NO PRODUCT LA NOUVELLE DONNE DE LA MONDIALISATION

Le mot d’ordre de la mondialisation, du point de vue économique, c’est désormais « *no translation, no product* » selon la formule de Suzanne Topping [2000 :111] : « pas de traduction, pas de produit ». La formule apparaîtra expéditive voire brutale à d’aucuns, mais elle reflète bel et bien la réalité du monde contemporain : d’un bout à l’autre de la planète, il est impensable qu’un produit, quel qu’il soit, soit lancé autrement qu’en « localisant » [FOLARON 2007] la campagne publicitaire qui l’accompagne sous toutes ses facettes jusques et y compris la documentation². Entendue en ce sens, la localisation signifie en effet [DUNNE 2006 : 4] : « *Ensemble des processus par lesquels des contenus et produits numériques développés en un lieu (compris en terme de zone géographique, de langue, de culture) sont adaptés pour leur vente et leur utilisation en un autre lieu. La localisation inclut dès lors (a) la traduction du contenu verbal selon les conventions textuelles et linguistiques de la langue réceptrice et (b) l’adaptation du contenu non verbal – depuis les couleurs, les icônes, le format bitmap jusqu’à l’emballage et autres contraintes formelles. Elle inclut aussi toutes les données et les paramètres utiles à la prise en compte des exigences culturelles, techniques et légales du lieu visé. En somme, la localisation ne porte pas tant sur des tâches spécifiques que sur des processus d’adaptation de produits* ». On peut cependant étendre la perspective au-delà : c’est vrai de l’espace numérique, mais également de tous les espaces économiques. Le cyberspace n’est en fait que la face émergée de l’iceberg : ce n’est pas parce que le

1 Voir dans cet ouvrage : Michaël Oustinoff, *L’anglais ne sera pas la lingua franca de l’internet*.

2 Voir dans cet ouvrage : Dwayne Bailey, *Localisation des logiciels : open source et multilinguisme numérique*.

cyberespace est devenu massivement multilingue que le multilinguisme tend à s'imposer de manière accrue à l'échelle planétaire, mais bien parce que le multilinguisme est une force économique majeure dans le monde d'aujourd'hui qu'il s'est imposé dans le cyberespace en tant que nouvel espace de l'économie de marché mondialisée.

Une étude récente réalisée par l'Université de Stanford au sujet de la Californie – qui, à elle seule, se classe parmi les dix premières économies mondiales – en constitue la parfaite illustration. L'argumentation d'ensemble est facile à résumer. En premier lieu, il est urgent d'investir dans les langues en raison des risques actuels de récession planétaire : « *Les réalités économiques, géographiques et sociales de la Californie font qu'il est urgent d'investir dans l'enseignement des langues étrangères maintenant. Alors que nous faisons face à une récession globale, la Californie (une des dix plus grandes économies au monde) connaît un déficit de 42 milliards de dollars, d'où le besoin crucial de trouver des solutions de long terme* » [STEELE 2009:1]. Le déplacement du centre de gravité économique vers l'Asie et l'Amérique du Sud doit, en effet, être pris en compte : « *Dans la première partie de ce rapport, nous soutiendrons que de telles solutions exigent que l'on se tourne résolument vers les autres parties du monde. Pour capitaliser sur la croissance des marchés des zones Asie-Pacifique et Amérique Centrale et Latine, la Californie dispose d'une position géographique idéale.* » (*ibid.*). Or la Californie peut faire appel à une population multilingue d'une très grande diversité : « *La Californie possède aussi une population dont l'exceptionnelle diversité accroît d'autant ses compétences linguistiques. C'est grâce aux dizaines de millions de personnes qui en Californie parlent des langues telles l'espagnol, le chinois, le coréen que nous avons une position nous permettant d'exceller dans le domaine du commerce, de l'économie et de la diplomatie* ». La conclusion s'impose : « *Accroître la valeur de notre force de travail en mettant l'accent sur l'enseignement des langues étrangères, c'est savoir utiliser pleinement à la fois notre position stratégique et nos ressources humaines afin de faire décoller l'économie* » (*ibid.*). L'étude n'hésite pas, à plus court terme, à parler d'un véritable « programme de relance » (*stimulus plan*) en des termes que ne désavouerait sans doute pas le Président Barack Obama :

Même dans une économie en déclin, il existe des moyens pour que la Californie connaisse une croissance de ses marchés et des emplois. L'heure est venue de tirer le meilleur parti de la puissance unique que

constitue le multilinguisme pour la Californie: une population importante, diverse, composée de plusieurs millions de personnes bilingues voire plurilingues en mesure de passer couramment d'une culture à l'autre, ainsi qu'une situation géographique idéale qui lui permet d'atteindre les marchés de l'autre côté du Pacifique ou en Amérique Latine ou Centrale. Face à la crise financière, investir dans l'enseignement des langues du monde et des compétences multi-culturelles devient une composante indispensable d'un plan de redressement économique efficace.» (ibid. : 3)

Néanmoins, il ne faut pas croire que les choses se mettront en place d'elles-mêmes, comme par enchantement. Certes, la demande en la matière est de plus en plus forte: «*La part croissance de la Californie dans le commerce international et la diversité des partenaires commerciaux attise une forte demande pour les Californiens qui savent appliquer leurs langues et leurs cultures à l'économie.» (ibid. : 4).* Mais encore faut-il savoir y répondre de manière adéquate: «*Les marchés internationaux sont difficiles, et même de grandes et riches entreprises s'y sont cassé les dents faute de compétences linguistiques et culturelles.» (ibid.).*

Les déboires de KFC (*Kentucky Fried Chicken*) en Chine en sont la démonstration éclatante. Dans un premier temps, le slogan de KFC «*finger lickin' good*» («*délicieux à s'en lécher les doigts*») avait été traduit littéralement en chinois pour aboutir au message aussi sinistre que cocasse: «*mangez vos doigts*» («*eat your fingers off*»), avec les résultats désastreux que l'on imagine aisément quand on cherche à s'implanter sur un marché local de cette importance. Les pertes financières d'une telle bourde s'évaluent à des millions de dollars en termes de parts de marché et d'image. Aujourd'hui, KFC est devenu la chaîne de restauration rapide numéro un en Chine. Mais il aura fallu faire machine arrière et réagir au plus vite: «*L'entreprise a dû produire en urgence un slogan qui fasse sens en Chine, remplacer la salade de chou par des mets traditionnels locaux, avant de devenir la première chaîne de Fast-Food américaine en Chine. Au début des années 2000, deux millions de Chinois mangent chez KFC chaque jour, et l'entreprise ouvre 250 nouveaux restaurants par an. Une victoire commerciale retentissante intimement liée à un savoir-faire d'ordre linguistique et culturel»(ibid. : 4).*

On pourrait multiplier les exemples, mais on l'aura compris, l'essentiel c'est qu'une telle vision des choses était tout bonnement impensable aux États-Unis il y a seulement une vingtaine d'années. Rappelons-nous: après

la chute du Mur de Berlin et l'effondrement de l'URSS, on nous expliquait que le monde serait désormais unipolaire, avec les USA pour seule hyperpuissance au centre de « l'espace monde » [DOLLFUSS 1994]. En vertu des effets du *soft power* [NYE 2004], les langues s'ordonnaient toutes pareillement autour de la langue « hypercentrale » [CALVET 2007]. L'économie des langues se résumait à celle de leur rapport au *Global English*, l'anglais « planétaire ». Un rapport par définition inégal : l'intérêt économique d'une *lingua franca* universelle n'était-il pas, justement, de se passer des autres langues, conçus comme autant de coûts supplémentaires ?

L'ÉCONOMIE DES LANGUES DANS UN MONDE MULTIPOLAIRE POST-AMÉRICAIN

L'exemple de la Californie, État d'une richesse multiculturelle exceptionnelle mais dont le potentiel économique correspondant n'a, pour l'instant, jamais véritablement été exploité est révélateur d'un paradoxe que Fareed Zakaria met ainsi en lumière : « *Dans plusieurs générations, quand les historiens écriront sur le temps présent, ils pourraient relever qu'au cours des premières décennies du XXI^e siècle, les États-Unis ont rempli avec succès leur grande mission historique – ils ont mondialisé le monde. Mais en chemin, pourraient-ils écrire, ils ont oublié de se mondialiser eux-mêmes* » [ZAKARIA 2008:95].

On pourrait y ajouter... l'Europe ! Certes, l'Union Européenne a pour fondement la défense de la pluralité linguistique, l'Europe des 27 comptant aujourd'hui 23 langues officielles. Elle entend promouvoir le plurilinguisme au sein de l'Union, comme le soulignait l'ancien Commissaire européen au multilinguisme, Leonard Orban : « *Les salariés devraient donc maîtriser, pour le plus grand bonheur de leurs employeurs, au moins trois langues : celle de leur pays d'origine, l'anglais bien sûr, et une troisième parmi les plus parlées de l'UE – l'allemand, le français, l'espagnol ou l'italien. Sans négliger le russe, l'arabe ou le chinois* » [RICARD 2007]. Néanmoins, une telle politique tarde à faire sentir ses effets au sein des entreprises, où le manque à gagner que constitue une maîtrise insuffisante des langues étrangères est criant, à commencer par la Grande-Bretagne, où il s'élèverait à 100 milliards d'euros par an :

M. Orban cite une étude réalisée fin 2006 par l'organisme d'expertise linguistique homologué par le gouvernement... britannique. D'après

cette enquête, 11 % des 2000 petites et moyennes entreprises (PME) interrogées dans l'Union européenne (UE) ont déjà perdu un contrat à l'exportation, faute de compétences linguistiques suffisantes. Soit un manque à gagner estimé à 100 milliards d'euros par an. « L'idée selon laquelle l'anglais est la langue universelle pêche par simplisme », observe l'étude. Privilégiée par les multinationales, la langue de Shakespeare serait certes utilisée en premier lieu pour négocier sur une vingtaine de marchés. Mais l'allemand permettrait d'exporter plus facilement vers quinze marchés, dont l'Allemagne et l'Autriche. Quant au français, il est employé sur huit marchés dont la France, la Belgique et le Luxembourg. [RICARD, IBID.]

En dépit des différences (effets de la mondialisation dans le premier cas, échanges au sein de l'Union Européenne dans l'autre), la similitude entre la Californie et la Grande-Bretagne est facile à établir : le décloisonnement des marchés entre États-nations, « *plus ou moins protégés jusque-là* » [DAVEL 2008:6] que ce soit à l'échelle locale, mondiale ou « *glocale* ». Mais cette explication, en l'occurrence, ne saurait suffire : ce qui est spectaculaire, c'est l'abandon de l'étalon « anglais » (de même qu'on a connu l'abandon de l'étalon or), ou, plus exactement, l'appel à d'autres étalons. Autrefois, ni la Grande-Bretagne (hyperpuissance du XIX^e siècle jusqu'à ce que les USA prennent la relève à partir de la Seconde guerre mondiale) ni les États-Unis ne s'inquiétaient guère de réciprocité en termes de langues, l'anglais s'imposant comme *lingua franca* de référence.

Aujourd'hui, la donne est radicalement différente, et ce pour trois raisons fondamentales. La première, c'est l'échec du modèle du tout-à-l'anglais, qui devait nous permettre de « communiquer » à l'échelle planétaire par l'intermédiaire d'une seule langue. Force est de constater que la maîtrise généralisée de cette langue est allée bien moins vite que le décloisonnement des marchés. La deuxième raison, liée directement à la précédente, réside dans la constatation que c'est dans la langue de l'Autre qu'il faut vendre, et non dans une langue étrangère, fût-ce l'anglais. La troisième raison est relativement nouvelle, et sans doute la plus déterminante à plus long terme : nous serons de plus en plus amenés à vivre dans un monde multipolaire, où les États-Unis ne seront plus le seul maître du jeu. Que les États-Unis soient en déclin [TODD 2002] ou simplement confrontés à ce que Fareed Zakaria appelle « *l'ascension des autres* » (notamment les BRIC : Brésil, Russie, Inde, Chine), peu importe : ces puissances dites

« émergentes » ne peuvent que prendre conscience de l'importance du *soft power* – et même du pouvoir tout court – que représentent leurs langues respectives. L'engouement actuel pour la langue chinoise, pourtant réputée d'une complication extrême, ne s'explique pas autrement : en tant que langue de la deuxième puissance mondiale, son apprentissage s'inscrit dans l'espérance d'un gain d'ordre économique, même si d'autres motivations ne sont pas à exclure (d'ordre culturel, notamment).

L'ÉMERGENCE D'UN NOUVEAU CHAMP DE L'ANALYSE ÉCONOMIQUE : LES LANGUES COMME VALEUR

Au vu de ce qui précède, on ne s'étonnera pas de ce que les économistes se soient très peu intéressés aux langues en tant que telles jusqu'à une époque très récente. Les premières études, en effet, remontent aux années 1960. Trois séries de raisons, semble-t-il, permettent de comprendre une telle désaffection. La première est à relier, dans le monde occidental, à la tradition gréco-romaine, héritée de l'Antiquité : pour les Grecs, il n'existait qu'une langue, la leur. Le monde grec est un monde fondamentalement monolingue. C'est ce modèle qu'ils ont transmis aux Romains, qui nous l'ont transmis à leur tour [CASSIN 2004] : les langues étrangères n'y jouent jamais, au mieux, qu'un rôle purement instrumental, autrement dit de second ordre par rapport aux processus plus élevés de l'esprit ou de l'activité humaine. Il faut attendre la fin du XIX^e siècle pour voir la langue devenir l'objet d'une science, en l'occurrence la linguistique.

La deuxième série de raisons viennent d'être évoquées : ce n'est que récemment qu'apparaissent le décloisonnement des marchés, la mondialisation, l'émergence d'un monde multipolaire, sans oublier les TIC qui constituent une véritable révolution en la matière. Pour des raisons aussi bien théoriques qu'empiriques, les langues ne pouvaient être considérées comme objet de réflexion véritable pour les économistes. Qui plus est, ces deux premières séries de raisons se combinant l'une à l'autre, les langues étaient essentiellement envisagées comme un coût (en tant qu'obstacle à l'accès direct aux « concepts » ou à l'agir sur le monde).

Mais il y a un troisième point de vue à considérer, d'ordre plus proprement méthodologique, qui permet de comprendre que « *l'économie de la langue reste un domaine de spécialisation assez marginal, en bordure de*

l'économie» [GRIN 2002 :14]. Deux éléments sont à prendre en compte: « 1. *l'économie de la langue est nécessairement une entreprise interdisciplinaire et [...] l'économie dite « orthodoxe », en tant que discipline, est notoirement fermée à l'interdisciplinarité*; 2. *la modélisation économique théorique et l'ensemble de concepts sur lesquels elle repose exigent le recours à des variables quantitatives, ou tout du moins à des variables qui se prêtent à une interprétation en termes de « plus » et de « moins ». Or l'étude des questions de langue demande habituellement que l'on tienne compte de variables qui se prêtent peu à l'interprétation quantitative, rendant ainsi moins pertinente la modélisation. Étant donné la popularité de la modélisation dans les départements d'économie au sein des universités, tout domaine de spécialisation qui ne se prête pas à l'application d'outils algébriques sera probablement négligé* » [GRIN *IBID.* :14-15].

Cela étant, on retiendra une distinction majeure dans la définition de la « valeur » d'une langue. On peut lui assigner, tout d'abord, bel et bien une valeur marchande: « *Les valeurs marchandes sont reflétées dans les prix ou dans un autre indicateur de ce type. Supposons, par exemple, que parler la langue X facilite la vente de biens au public parlant la langue X et permet en conséquence de réaliser des bénéfices plus importants ou bien qu'un salarié parlant la langue X gagne davantage, toutes choses égales par ailleurs, car il connaît la langue X; dans ce cas, la langue X a une valeur marchande* » [GRIN *IBID.* :21].

Deuxième facette fondamentale de la valeur, la valeur non marchande: « *Par exemple, connaître la langue X donne accès à la culture en langue X, facilite les contacts sociaux avec les membres de la communauté parlant la langue X, etc. Cette valeur n'est généralement pas reflétée dans les prix marchands [...]* » [GRIN *IBID.* :21].

Par ailleurs, la valeur est à considérer sur deux plans: à l'échelle individuelle, ou « privée », et à l'échelle collective, ou « sociale ». Par conséquent, on se retrouve avec un tableau à quatre entrées.

Si l'on se met à combiner ces quatre dimensions (valeur marchande privée vs sociale, valeur non marchande privée vs sociale) avec ce qui précède, on voit toute la complexité des cas de figure à considérer. D'où, notamment, en l'état actuel des connaissances, la difficulté à évaluer exactement la valeur d'une « langue »: « *En bref, il n'est pas possible, pour le moment, de véritablement calculer 1. la « valeur » d'une langue; 2. la « valeur » d'un*

environnement linguistique par rapport à un autre; 3. les « avantages » (marchands et non marchands) qui peuvent être attendus d'une politique particulière; 4. beaucoup des coûts, directs et indirects, associés à une telle initiative» [GRIN IBID. :21].

Ce qui n'infirme pas pour autant toute analyse économique, bien au contraire. En particulier, François Grin fait apparaître un inconvénient majeur du tout-à-l'anglais, que David Graddol [GRADDOL 1997 ET 2006] signale également, lorsqu'il entre en concurrence avec d'autres langues : *« On peut avancer à juste titre que même si l'anglais est rentable, favoriser l'apprentissage de cette langue rendra en conséquence sa connaissance plus banale et les avantages de salaire dont bénéficient les anglophones seront peu à peu érodés [...] Cette évolution affectera vraisemblablement la plupart des pays dans lesquels l'anglais est (de plus en plus) enseigné comme seconde langue et jugée indispensable à l'activité économique. Autrement dit, des indices convergents amènent à penser que la connaissance d'autres langues deviendra peu à peu relativement plus rentable à mesure que l'anglais se généralise, ce qui implique que les politiques d'enseignement des langues ne devraient pas être axées sur l'anglais seul en tant que deuxième langue, mais sur d'autres langues également » [GRIN IBID. :34].*

Autrement dit, les premières victimes collatérales de cette nouvelle donne de la mondialisation sont les monologues anglophones, mais pas uniquement. Ce qui est vrai pour eux pourrait donc fort bien, à terme, valoir pour les autres. L'analyse suivante devient alors généralisable : *« Le coût de l'apprentissage de l'anglais pour les non-anglophones continue de baisser, alors que le coût d'apprentissage des autres langues pour les anglophones (mais aussi pour tous les autres) est en augmentation. Une des conséquences de l'extension universelle de la lingua franca serait alors que les anglophones vont entrer en compétition sur leur propre marché du travail avec le reste du monde, sans avoir réellement d'accès aux marchés du travail pour lesquels une autre langue est nécessaire » [VAN PARIJS 2004].* Dans un monde post-américain, la valeur marginale (au sens économique du terme) de l'anglais décline au fur et à mesure que celle des autres langues augmente d'autant.

Vu l'importance de l'industrie des langues [RINSCHÉ 2009] et son taux de croissance qu'on prévoit se maintenir durablement à 10 %, il n'est pas besoin d'être grand clerc pour comprendre les enjeux économiques des langues dans le monde contemporain – enjeux qui se situent aujourd'hui à l'échelle des États et non plus seulement à l'échelle individuelle. C'est par

conséquent aussi en termes politiques que se pose désormais la question des langues, voire en termes de gouvernance internationale à l'heure de la mondialisation.

CONCLUSION

Le tout-à-l'anglais était censé représenter la solution la plus pratique et, par voie de conséquence, la plus économique [GRIN 2004]. On s'aperçoit aujourd'hui qu'il n'en est rien : loin de constituer un avantage, ce modèle représente en réalité un manque à gagner considérable, ce dont prennent conscience les pays anglophones eux-mêmes, à commencer par la Grande-Bretagne et les États-Unis. Autrefois considérée comme une question de second ordre, le potentiel linguistique d'une nation constitue désormais un atout stratégique majeur à l'heure de la mondialisation de l'économie. La formule de Suzanne Topping « *No translation, no product* » mérite sans doute d'être nuancée, mais elle n'en traduit pas moins un changement de paradigme majeur, qui ne cantonne pas les langues à leur valeur non marchande mais leur octroie une valeur marchande majeure, non seulement en termes de coûts mais également en termes d'investissement rentable [GRIN 2009].

L'économie des langues ne saurait se réduire à sa dimension linguistique, au sens traditionnel du terme. Il faut y ajouter la dimension culturelle. C'est ce qu'ont compris les écoles de management anglophones les plus prestigieuses au début des années 2000 [EARLEY 2003 ; EARLEY 2004 ; CHUA 2009], en mettant l'accent sur l'« intelligence culturelle » (*Cultural Intelligence*, ou CQ pour « Cultural Quotient ») dont les *Cultural Studies* ou l'Unesco avaient depuis longtemps reconnu l'importance :

[...] le village global est une mosaïque multiculturelle. Des pays différents ont des cultures différentes. Il en résulte la nécessité pour les entreprises et les gestionnaires d'user de prudence en essayant de mettre en place the best way à l'échelle de la planète entière. Quand de telles différences ne sont pas prises en compte, les entreprises adoptent des stratégies et des politiques inappropriées. Quant aux gestionnaires, ils prennent les mauvaises décisions, se heurtent aux cultures locales, éprouvent un grand désajustement culturel et ne parviennent pas à mener à bien leurs initiatives. [REGO 2009:34]

Il est malheureusement moins sûr que les États soient aujourd'hui en train de prendre les mesures qui s'imposent : le retard en la matière est considérable, puisque les politiques linguistiques qu'ils ont mis en place depuis la Seconde guerre mondiale ont fait de l'enseignement de l'anglais l'objectif numéro 1. Par ailleurs, les coupes budgétaires touchent actuellement de plein fouet le service public, celles-ci s'abattant en priorité sur les autres langues que l'anglais. C'est courir un risque majeur, celui d'aggraver ce que l'on appellera la « fracture linguistique » sur le modèle de la fracture numérique : en s'en remettant à la « main invisible » du marché, on accroît les disparités de l'accès à un plurilinguisme véritablement opérationnel, qui restera confiné aux élites. Or, dans une économie mondialisée et à l'heure de l'enseignement de masse, former seulement les élites ne suffit pas [OUSTINOFF 2011] : c'est un modèle obsolète aux conséquences sociales et économiques à long terme que l'on sait aujourd'hui désastreuses. Telle est la leçon que l'économie des langues nous apprend : les décideurs et les politiques sauront-ils l'entendre ?

Il reste néanmoins un domaine prometteur qui incite à l'optimisme : celui de l'espace numérique, grâce au déploiement des TIC, qui permettent de réduire les coûts de manière drastique, voire de les annuler. L'exemple de Wikipédia en est une des illustrations les plus éclatantes : il suffit de disposer d'un ordinateur et d'une connexion internet pour disposer des ressources mises en ligne gratuitement et ce, dans les langues les plus diverses. C'est d'ailleurs ainsi que Wikipédia se définit en ligne : « Wikipédia est un projet d'encyclopédie collective établie sur internet, universelle, multilingue et fonctionnant sur le principe du wiki. Wikipédia a pour objectif d'offrir un contenu librement réutilisable, objectif et vérifiable, que chacun peut modifier et améliorer. » C'est un modèle qui a fait des émules, comme en témoigne l'émergence de sites tels que Global Voices, où l'on trouve sur le versant en français de son site l'explication suivante : « Global Voices Online est une organisation à but non lucratif et un réseau de plus de 500 blogueurs et traducteurs dans le monde entier, fondés au Berkman Center for Internet and Society de la faculté de droit de Harvard. Ils proposent des revues de blogs du monde entier, traduites en 18 langues, en accordant une attention toute particulière aux voix absentes des médias traditionnels. » De telles initiatives sont appelées à se multiplier, et l'on voit tout l'intérêt qu'elles représentent. C'est, ni plus ni moins, mettre à la disposition du plus grand nombre les innombrables applications rendues possibles par les TIC en permettant à chacun d'y avoir accès dans sa langue (sans oublier

le plan de l'enseignement des langues et de l'autodidaxie). Néanmoins, même la gratuité a un coût, comme n'ont aucun mal à nous l'apprendre les économistes, pour qui « un repas gratuit, ça n'existe pas » (« *There ain't no such thing as a free lunch* »). C'est pourquoi, là encore, s'en remettre à la seule logique des marchés serait la pire des solutions.

BIBLIOGRAPHIE

[BRITISH ACADEMY 2009] The British Academy. 2009. *Language Matters. Position Paper*. <http://www.britac.ac.uk>

[CALVET 2007] Calvet, L.-J. 2007. La traduction au filtre de la mondialisation, In : *Traduction et mondialisation*, ed Oustinoff, M. and Nowicki, J. *Hermès* 49. Paris : CNRS Éditions.

[CASSIN 2004] Cassin, B. 2004. *Vocabulaire européen des philosophies. Dictionnaire des intraduisibles*. Paris : Le Robert/Le Seuil.

[CRONIN 2003] Cronin, M. 2003. *Translation and Globalization*. London : Routledge.

[CHUA 2009] Chua, R.Y.J. and Morris, M. W. 2009. Innovation Communication in Multicultural Networks : Deficits in Inter-cultural Capability and Affect-based Trust as Barriers to New Idea Sharing in Inter-cultural Relationships. *Harvard Business School Working Paper 09 (130)*, May.

[DAVEL 2008] Davel, E., Dupuis, J.-P., and Chanlat, J.-F. 2008. *Gestion en contexte interculturel. Approches, problématiques, pratiques et plongées*. Quebec : Presses de l'Université Laval.

[DOLLFUSS 1994] Dollfuss, O. 1994. *L'espace monde*. Paris : Economica.

[DUNNE 2006] Dunne, K. (ed). 2006. *Perspectives on Localisation. ATA Scholarly Monograph Series 8*, Amsterdam and Philadelphia : John Benjamins Publishing.

[DWYER 2010] Dwyer, T. 2010. Traducteurs, associations professionnelles et marché : approches empiriques, In : Oustinoff, M., Nowicki, J. and Machado da Silva, J. (eds). 2010. *Traduction et mondialisation, Volume 2. Hermès* 56. Paris : CNRS Éditions.

[EARLEY 2003] Earley, C. and Ang, S. 2003. *Cultural Intelligence: Individual Interactions across Cultures* Palo Alto : Stanford University Press.

[EARLEY 2004] Earley, C. and Mosakowski, E. 2004. Cultural Intelligence. *Harvard Business Review* Oct.

[FOLARON 2007] Folaron, D. and Gambier, Y. 2007. La localisation : un enjeu de la mondialisation, in Oustinoff, M. and Nowicki, J. (eds). 2007. *Traduction et mondialisation. Hermès* 49. Paris : CNRS Éditions.

[GRADDOL 1997] Graddol, D. 1997. *The Future of English? A Guide to Forecasting the Popularity of the English Language in the 21st Century*. The British Council & The British Company (UK) Ltd. (new ed. 2000). <http://www.britishcouncil.org/learning-elt-future.pdf>

[GRADDOL 2006] Graddol, D. 2006. *English Next. Why Global English may Mean the End of "English as a Foreign Language*. The British Council & The British Company (UK) Ltd, 2006 (new ed., 2007). <http://www.britishcouncil.org/learning-research-english-next.pdf>

[GRIN 2002] Grin, F. 2002. *L'économie de la langue et de l'éducation dans la politique de l'enseignement des langues*. Strasbourg : Language Policy Division, Council of Europe.

[GRIN 2004] Grin, F. 2004. Coûts et justice linguistique dans l'élargissement de l'Union européenne. *Panoramiques*, n° 69, 4e trimestre 2004, 97-104.

[GRIN 2009] Grin, F. 2009. *The Economic Value of Multilingualism: Private, Social and Macroeconomic Perspectives*. Brussels : Directorate-General for Translation, European Commission. 27 Nov. http://webcast.ec.europa.eu/eutv/portal/_v_fl_300_fr/player/index_player.html?id=8095&pId=8093

[NYE 1997] Nye, J. 1997. *Soft Power: The Means to Success in World Affairs*. New York : Public Affairs.

[OUSTINOFF 2011] Oustinoff, M. 2011. *Traduire et communiquer à l'heure de la mondialisation*. Paris : CNRS Éditions.

[REGO 2009] Rego, A. and Pina e Cunha, M. 2009. *Manuel de gestão transcultural de recursos humanos* (Handbook for Transcultural Management and Human Resources). Lisbonne : Editora RH.

[RINSCHÉ 2009] Rinsché, A. and Portera-Zanotti, N. 2009. *The Size of the Language Industry in the EU*. Brussels : Direction générale de la traduction, Commission européenne.

[RICARD 2007] Ricard, P. 2007. Une étude britannique prône le multilinguisme en affaires. *Le Monde*, 25 Sept.

[STEELE 2009] Steele, T., Oishi, L., O'Connor, K., and Silva, D. M. 2009. *Learning World Languages and Cultures in California : A Stimulus for Academic and Economic Success*. Stanford : Stanford University School of Education.

[TODD 2002] Todd, E. 2002. *Après l'empire: Essai sur la décomposition du système américain*, Paris : Gallimard.

[TOPPING 2000] Topping, S. 2000. Shortening the Translation Cycle at Eastman Kodak, in Sprung, R. C. (ed). 2000. *Translating into Success: Cutting-Edge Strategies for Going Multilingual in a Global Age*, Amsterdam and Philadelphia : John Benjamins.

[VAN PARIJS 2004] Van Parijs, P. 2004. Europe's Linguistic Challenge, *Ce, European Journal of Sociology*, 45 (1).

[ZAKARIA 2008] Zakaria, F. 2008. *The Post-American World*. New York : Norton.

DANIEL PRADO
& DANIEL PIMIENTA

QUELLES POLITIQUES PUBLIQUES POUR LES LANGUES DANS LE CYBERESPACE ?

De nombreuses recommandations et déclarations émanant de sommets, organisations internationales, rencontres, etc. suggèrent de multiples actions en matière de présence des langues dans le cyberespace afin de promouvoir la diversité linguistique dans la Société des savoirs. Toute politique d'aménagement linguistique numérique doit reposer sur une évaluation précise des langues dans le cyberespace.

Article original en français.



DANIEL PRADO est l'ancien directeur du Département terminologie et industrie de la langue de l'Union Latine, organisation intergouvernementale de promotion des langues et des cultures latines. Il est aujourd'hui Secrétaire exécutif du Réseau mondial pour la diversité linguistique, Maaya.



DANIEL PIMIENTA, né à Casablanca, a étudié les Mathématiques appliquées et l'informatique à l'Université de Nice. Il rejoint IBM France en 1975 comme architecte de systèmes de télécommunications. En 1988, il devient conseiller scientifique de l'Union Latine, à Saint-Domingue, et dirige un projet de réseau régional. En 1993, il fonde l'Association Réseaux & Développement (Funredes). En 2008, il reçoit le Namur Award (IFIP WG9.2) pour l'ensemble de ses actions en vue d'une vision holistique de l'impact social des TIC.

DANIEL PRADO
& DANIEL PIMIENTA

QUELLES
POSITIONS
PUBLIQUES
POUR LES
LANGUES DANS LE
CYBERSPACE ?

Les langues et la culture prennent une place de plus en plus importante dans les projets politiques contemporains, au niveau national, régional et international.

Notre monde a beaucoup changé depuis le partage de Yalta au lendemain de la Seconde guerre mondiale, quand deux façons différentes de voir l'avenir se confrontaient et quand les relations internationales se basaient sur des critères militaires, politiques et économiques. Dans ce schéma bipolaire, la valorisation du fait culturel comme base de la société était complètement ignorée. Nous constatons également que le monde n'est déjà plus celui que nous avons connu à partir de la fin des années quatre-vingt, qui voyait s'imposer, sous l'intitulé trompeur de « fin de l'histoire », un modèle politique et économique transmettant une culture unique et tendant à l'hégémonie.

Au XXI^e siècle le monde est devenu multipolaire. Les alliances supranationales basées dans le respect du fait culturel se sont renforcées, des alliances entre régions ou entre différentes Nations partageant une même langue ou une même culture naissent tous les jours. Si les enjeux politiques et économiques sont toujours présents dans les relations internationales, les aspects culturels prennent une place de plus en plus grande. La religion, l'ethnie, la coutume, la langue participent aujourd'hui à la définition des politiques internationales, même si parfois, hélas, elles sont facteur de discorde. Le fait culturel est de plus en plus partie prenante des éléments du développement durable et de la croissance équitable, facilitant l'harmonie des peuples et le respect de leur dignité.

La langue est dorénavant présente dans les enjeux internationaux, tant politiques que commerciaux. Au côté de l'anglais, les instances internationales et régionales, reconnaissent d'autres grandes langues de

communication co-officielles : français, espagnol, russe, chinois, arabe aux Nations Unies ; 22 autres langues à l'Union Européenne ; portugais, français, arabe et swahili à l'Union africaine ; français, espagnol et portugais dans les enceintes du continent américain, etc. Celles-ci réclament un respect d'équilibre dans leur traitement.

D'autres langues qui ne sont pas officielles (portugais à l'ONU, langues régionales auprès de l'Union Européenne, guarani en Amérique du Sud, etc.) réclament leur droit à la co-officialité. Des initiatives provenant de divers organismes (Conseil de l'Europe, Linguapax, Unesco, etc.) tendent à donner un rôle de plus en plus croissant aux langues qui n'ont pas de statut d'officialité nationale ou régionale.

Cette régionalisation des rapports rend plus fortes et plus présentes des langues qui avaient perdu du terrain il y a quelques décennies et qui aujourd'hui reviennent dans l'enseignement et le commerce. La mondialisation économique s'accompagne d'une multiplication des langues utilisées par les industries. Selon une enquête assez médiatisée par l'administration étasunienne qui a été réalisée dans le but de favoriser les industries linguistiques¹, 60 % des consommateurs des pays développés n'achètent jamais un produit qui n'est pas décrit dans leur langue. Les entreprises ont tardé à le comprendre mais elles localisent de plus en plus², même si les administrations nationales et internationales, souffrant d'inerties bureaucratiques, prennent un peu plus de temps à réagir.

L'adoption de la « *Convention sur la protection et la promotion de la diversité des expressions culturelles* »³, promue par l'Unesco, a été le symbole même du changement de la place de la culture dans les relations internationales. Au-delà d'une simple valeur économique, portée par ses industries spécialisées, la culture est surtout une condition essentielle de l'être humain et le meilleur moteur pour un développement respectueux du futur de la planète.

Cette Convention a ouvert une brèche que les militants du multilinguisme et du respect de la diversité linguistique ont bien saisie, car il était temps de considérer le fait linguistique comme intimement associée à la libre

1 *Report on Global Consumer Online Buying Preferences, Showing the Impact of Language, Nationality, and Brand Recognition*, Common Sense Advisory, 2006.

2 Voir dans cet ouvrage : Michaël Oustinoff, *L'Économie des langues*.

3 Adoptée par l'Unesco le 20 octobre 2005 <http://www.unesco.org/new/fr/culture/themes/cultural-diversity/diversity-of-cultural-expressions/the-convention/convention-text/>

expression, à l'épanouissement, à l'égalité des chances, à la promotion de l'entente entre les peuples sur des bases justes et équilibrés.

La langue a certes une place implicite dans cette Convention, mais l'implicite n'est pas évident pour tout le monde. Les Objectifs du Millénaire pour le Développement⁴, oubliant d'intégrer la culture en tant qu'objectif unique ont également omis de parler de la langue des individus. Certes, la *Convention sur le patrimoine immatériel*⁵ lui fait une mention attentive et, ensuite, tant la *Recommandation sur le multilinguisme dans le cyberspace*⁶ que la *Déclaration de principes du Sommet mondial sur la Société de l'information*⁷ donnent corps à l'idée du respect de la diversité linguistique et du multilinguisme. Mais encore faut-il que ces instruments soient vraiment suivis de réalisations concrètes. Et nous savons que nous sommes loin d'obtenir pour tous les individus l'opportunité de se développer librement dans leur langue.

Devant l'absence de conventions internationales concernant les langues en cours d'existence, devant l'absence d'indicateurs fiables sur leur place et impact dans le développement mondial, devant la disparition probable de près de la moitié d'entre elles, devant le nombre encore trop important d'injustices liées au fait de parler une langue non reconnue, des mesures s'imposent pourtant. Elles concernent le domaine des indicateurs, des politiques publiques, la promotion du multilinguisme et de la diversité linguistique, sans oublier les instruments juridiques.

S'il est vrai que moins d'un individu sur trois dans le monde dispose aujourd'hui d'un accès à internet⁸, l'évolution du réseau est permanente et tend vers l'universalisation. Nous savons également que le cyberspace et les technologies associées tendent à remplacer progressivement nos anciennes manières de communiquer, de nous exprimer, de transmettre des informations, de partager des connaissances, de nous mettre en

4 Voir dans cet ouvrage : Adama Samassékou, *Multilinguisme, Objectifs du Millénaire pour le Développement et cyberspace*.

5 Adoptée par l'Unesco le 17 octobre 2003.

<http://www.unesco.org/culture/ich/index.php?lg=fr&pg=00006>

6 Recommandation concernant la promotion et l'usage du multilinguisme et l'accès universel au cyberspace, adoptée par l'Unesco le 15 octobre 2003.

<http://www.unesco.org/new/fr/communication-and-information/access-to-knowledge/linguistic-diversity-and-multilingualism-on-internet/normative-instruments/recommendation/>

7 Adoptée à Genève en décembre 2003.

<http://www.itu.int/wsis/docs/geneva/official/dop-fr.html>

8 <http://www.internetworldstats.com/stats.htm>

relation avec les autres. Les langues qui ne pourront pas faire circuler ces informations, connaissances ou dialogues, risquent de perdre de la valeur aux yeux mêmes de leurs locuteurs. Or les phénomènes migratoires et d'urbanisation ainsi que l'accès universel au réseau et aux médias provoquent des tensions inter-linguistiques dans lesquelles seules les langues bien valorisées par leurs locuteurs pourront survivre.

Malgré une avancée certaine du multilinguisme sur internet ces derniers temps, seule une petite poignée des langues du monde jouit d'une présence notable sur la Toile et l'anglais demeure la langue la plus utilisée, même si sa présence relative est en constante diminution et pourrait tendre vers le chiffre de 30 % des pages web⁹, qui correspond au pourcentage d'utilisateurs d'internet maîtrisant la langue anglaise¹⁰. Jean-Paul Corbeil nous l'annonçait en 2000 en affirmant qu'« à très brève échéance la présence de l'anglais devrait diminuer à plus ou moins 40 % lorsque des sites seront créés dans divers pays, au fur et à mesure qu'ils se brancheront sur le réseau »¹¹.

La réalité est que sur l'ensemble des langues présentes dans la planète, il n'y aurait qu'un faible nombre qui bénéficieraient d'institutions veillant à leur protection, essor ou équipement, et à peine 500 d'entre elles¹² auraient une présence écrite dans le cyberspace (par contre, il y en aurait beaucoup plus si on compte les ressources audiovisuelles). Le développement d'une langue ne passe pas seulement par la volonté et la capacité d'une institution à la rendre apte à toute utilisation; c'est souvent la volonté d'individus, du corps enseignant ou de petites structures associatives travaillant sur le terrain qui participent à cette évolution¹³.

9 À la date de la rédaction de cet article, le chiffre estimé est de 26,8% d'utilisateurs de l'internet de langue anglaise, talonné par le chinois avec 24,6% puis l'espagnol avec 7,8% <http://www.internetworldstats.com/stats7.htm>

10 Voir dans cet ouvrage : Michaël Oustinoff, *L'anglais ne sera pas la lingua franca de l'internet*.

11 Corbeil Jean-Claude, « I comme informatique, industries de la langue et Internet », In B. Cerquiglini et al., *Tu parles ! ?, le français dans tous ses états*, Paris, Flammarion, 2000, p. 129.

12 Il s'agit du nombre de langues identifiées par Unicode comme ayant une représentation numérique http://unicode.org/repos/cldr-tmp/trunk/diff/supplemental/languages_and_scripts.html

Parmi celles-ci on estime à 300 celles qui sont effectivement utilisées sur internet. L'application de loin la plus multilingue est Wikipédia avec 269 langues différentes en 2011 <http://stats.wikimedia.org/EN/Sitemap.htm>

13 Voir dans cet ouvrage : Evgeny Kuzmin, *Politiques linguistiques pour contrer la marginalisation des langues*.

Que ce soit en s'appuyant sur une institution solide ou sur des initiatives plus légères, il n'y a ni règles, ni science, ni praxis exemplaires en matière de politiques linguistiques. Si de nombreux ouvrages leurs sont consacrés, si de nombreux colloques leurs sont dédiés (parfois de manière explicite, mais le plus souvent, implicite), les politiques linguistiques (ou l'aménagement linguistique) ne sont objet d'étude que depuis quelques dizaines d'années et, du fait de la disparité de situations sociolinguistiques, elles ne peuvent orienter efficacement les praticiens que pour les langues les mieux étudiées (notamment les langues européennes et quelques langues asiatiques).

Ainsi, il est difficile de trouver deux modèles de politiques linguistiques concordantes selon le pays ou le territoire où cette politique est appliquée. Sous l'appellation « politiques linguistiques » on trouve un nombre assez hétérogène de mesures qui peuvent aller depuis des politiques de promotion, revitalisation ou protection d'une ou plusieurs langues jusqu'aux politiques d'éradication de celles-ci! Nous trouvons des politiques très élaborées et explicites tendant à rendre vivante une langue devenue minoritaire ou peu utilisée (Québec, Catalogne, Israël), comme des politiques de mise en valeur des langues autochtones (Mexique, Bolivie, Paraguay, Mali, Bénin, etc.), ou des politiques d'équité des citoyens vis-à-vis de leur État (Suisse, Luxembourg, Aruba, etc.). Il existe également et malheureusement des politiques bien moins louables d'hégémonie linguistique donnant priorité absolue à la langue d'État et bannissant l'usage des autres.

Une politique linguistique¹⁴, par exemple, pourra intervenir dans l'adoption d'un système d'écriture ou des règles grammaticales et orthographiques, par l'enrichissement du vocabulaire, par la fixation du statut d'une ou plusieurs langues (officialisation, langue de l'administration et de la justice, langue d'enseignement, langue régionale ou nationale, etc.), par son enseignement au niveau international, etc.

Pour toute politique linguistique, qu'elle prône une hégémonie linguistique ou, au contraire, qu'elle favorise une équité de droits des citoyens en permettant à ceux-ci de participer pleinement dans la société grâce à l'utilisation de leur langue première, les indicateurs sont un outil d'orientation fondamental. Ils permettent de diagnostiquer objectivement la

14 Voir sur ce sujet la page de Leclerc (CIRAL) sur les politiques linguistiques.
http://www.tfq.ulaval.ca/axl/monde/index_politique-Ing.htm

situation, d'en mesurer les évolutions et d'évaluer les effets des politiques conduites.

Or, bien que de nombreuses entités nationales ou internationales (Unesco, La Francophonie, Union Latine, SIL, Université de Laval, British Council, CPLP, etc.) proposent des statistiques, des suivis, des sondages, des études, et autres, et visent à fixer des paramètres plus ou moins stables pour l'observation d'une ou plusieurs langues, les indicateurs qui en sont issus diffèrent selon l'institut d'origine donnant des contradictions flagrantes et permettant des effets d'annonce qui finalement ne font parfois que nourrir une plus grande incertitude au lieu de clarifier le paysage.

Depuis la simple question de savoir combien de langues existent dans le monde (qui soulève celle plus épineuse de la frontière entre parlers), à celle des méthodes de mesure du nombre de locuteurs (notamment en fonction de leur aptitude ou niveau de maîtrise de cette langue), d'autres interrogations se font jour autour de l'élaboration d'indicateurs, comme la taille des corpus disponibles, l'utilisation réelle dans la vie quotidienne, dans l'éducation, dans l'administration, dans la santé, dans les médias, dans l'information scientifique et technique, etc.

De nouveaux paramètres sont venus alimenter les arguments permettant de donner plus de place à une langue par rapport à d'autres ou à équilibrer la présence entre langues : à savoir combien « pèse » une langue ou combien « vaut » une langue. Différents travaux réalisés par Grin¹⁵, par Graddol¹⁶, par López Delgado¹⁷, par Esperança¹⁸ en matière de valeur d'une langue (en l'occurrence français, anglais, espagnol et portugais)

15 Grin, F., *Compétences et récompenses. La valeur des langues en Suisse*, Fribourg, Éditions Universitaires Fribourg, 1999.

Grin, F., «English as economic value: facts and fallacies», *World Englishes*, n°20, 2001.

16 Graddol, D. *English Next. Why Global English may Mean the End of "English as a Foreign Language"*. The British Council & The British Company (UK) Ltd. 2006 (nvlle éd.. 2007). <http://www.britishcouncil.org/learning-research-english-next.pdf>

17 García Delgado, J.L. et al, *Economía del español. Una introducción* (2ª edición ampliada), Madrid, Editorial Ariel, 2008.

18 Esperança, José Paulo. *O Valor Económico da Língua Portuguesa Pode Ser Potenciado*. <http://www.portalingua.info/fr/actualites/article/valor-economica-portugues/>

ainsi que ceux menés par Calvet¹⁹ et d'autres²⁰ sur le poids des langues, essaient de donner des paramètres vérifiables permettant de décider de l'opportunité d'apprendre ou d'enseigner une langue, de la nécessité de favoriser la présence d'une langue dans des secteurs de la société où elle est moins présente, ou tout simplement pour se positionner sur le marché du travail ou mieux vendre un produit.

Or, seules les populations conscientes de cette opportunité ou défi, verront leurs langues s'épanouir, les autres risquent d'assister à leur mort inéluctable dans une planète mondialisée où le numérique s'impose graduellement.

Pour pouvoir proposer des actions de réajustement allant dans le sens du respect de la diversité linguistique, nous manquons d'une vision réelle de la situation des langues. Les statistiques sont incomplètes, les indicateurs faussés, les études biaisées et surtout, les études métriques sur les langues, balbutiantes.

L'observation de l'évolution des langues dans le cyberspace n'échappe pas à cette règle générale, bien au contraire, et permet de mettre en lumière les lacunes existantes en la matière. C'est précisément l'émergence même de l'internet qui pose aux langues la question nouvelle « d'opportunité » et « d'enjeu ». En effet, le cyberspace est un vrai défi pour toute langue qui se voit confrontée à une compétition nouvelle pour fournir l'accès à un vaste corpus d'informations, à défaut de quoi ses propres locuteurs vont l'abandonner progressivement en favorisant celle qui à leurs yeux serait plus « prestigieuse », ou en tout cas apportant plus d'informations dans le domaine qui les concerne. Mais c'est aussi une opportunité car ce média offre une capacité d'expression plus facile et moins onéreuse que les médias traditionnels (notamment l'édition papier ou l'audiovisuel hertzien) et donc, peut devenir le moyen idéal pour permettre la résurgence de cette langue.

19 Calvet, Louis-Jean (2002). *Le marché aux langues*. Paris : Plon.

On pourra utiliser également du même auteur un outil d'évaluation du poids des langues en fonction de plusieurs critères que l'on peut pondérer :

Alain Calvet et Louis-Jean Calvet, *Baromètre Calvet des langues du monde*.

<http://www.portalingua.info/fr/poids-des-langues/>

20 Maurais, Jacques et Morris, Michael A. (eds.) (2003). *Languages in a globalising world*. Cambridge : Cambridge University Press ; Wallraf, Barbara. "What global language". *The Atlantic Monthly* 286 : 5 (2.000) : pp. 52-66. ; Weber, George. "Top Languages". *Language Today*. Diciembre (1997) : 12-18.

Internet, un nouvel espoir pour les langues? Probablement, car le cyberspace permet précisément d'ouvrir la porte à des formes d'expressions auxquelles les circuits de l'édition traditionnelle ne s'intéressent pas. Les publications scientifiques en d'autres langues que l'anglais peuvent par exemple se maintenir et trouver une place, certes modeste, grâce à la facilité et au faible coût de publication sur la Toile, alors que des revues équivalentes publiées par des éditeurs traditionnels ne concerneraient qu'un nombre minime de lecteurs, un risque que peu d'éditeurs sont prêts à prendre.

Internet a permis, sans doute, à des minorités absentes de l'édition traditionnelle de s'exprimer, mais gardons-nous de penser que le tour est déjà joué. La relation entre l'accès à internet et la diversité linguistique mondiale reste inversement proportionnelle, comme nous le démontrions en 2008²¹; la fracture linguistique se calque pour l'instant sur la fracture numérique.

À notre connaissance, aucune étude globale ne nous donne un aperçu réel de la place des langues au niveau mondial sur la Toile, les réseaux sociaux, la messagerie, le clavardage, etc. Nous constatons cependant, à travers diverses études parcellaires, que la production des réseaux sociaux est en franche accélération par rapport à la production de pages web traditionnelles, même si le contenu est souvent éphémère²². Les études menées par Semiocast²³ en 2010 sur Twitter, par exemple, montraient que le malais et le portugais étaient bien plus utilisés que l'espagnol, l'allemand, le russe et l'italien, pourtant langues bien plus présentes sur le web traditionnel et disposant de politiques de traduction et de numérisation d'œuvres bien plus solides. Cette étude n'a pas été renouvelée depuis, mais les langues parlées en Indonésie, par exemple, pourraient être encore plus présentes aujourd'hui, car ce pays a la troisième plus importante

21 Daniel Prado, «Languages and cyberspace: Analysis of the General Context and the Importance of Multilingualism in Cyberspace» In: *Proceedings of the International Conference Linguistic and Cultural Diversity in Cyberspace*, (Yakutsk, Russian Federation, 2-4 July, 2008). http://www.ifapcom.ru/files/Documents/multiling_eng.pdf

22 Le site observatoire Portalingua <<http://www.portalingua.info>>, créé par l'Union Latine tente de répondre à ce problème en compilant et en mettant en parallèle les études et statistiques existantes sur la présence des langues dans différentes sphères de la société de la connaissance.

23 http://semiocast.com/static/downloads/Semiocast_Half_of_messages_on_Twitter_are_not_in_English_20100224_fr.pdf

«génération 140»²⁴ de la planète. On assiste à des phénomènes semblables sur Facebook. Un rapport émanant de publicitaires (à prendre donc avec toutes les précautions d'usages pour ce type de recensement) décrit par exemple une augmentation de 175 % des arabophones sur ce réseau social, alors même que la progression des anglophones stagne (augmentation de 45 % en 2011)²⁵. Ce qui n'est pas sans conséquences sur les changements sociaux, politiques et générationnels dans les parties du monde concernées²⁶.

C'est pour cela qu'il est important d'insister sur les aspects tant d'infrastructure que d'appropriation des technologies et de production de contenus (texte mais aussi multimédia). La promotion de l'usage de la langue dans les entités régionales ou nationales concernées doit se faire à tous les niveaux : éducatif, administratif, scientifique et technique, voire loisir. Mais elle doit se faire à la base, au niveau de l'accès des langues aux technologies et sur la base de données fiables.

Or, les études documentées et systématiques en matière d'observation des langues sur la Toile sont rares. On ne repère véritablement que l'initiative de Funredes et l'Union Latine menée de 1998 à 2007²⁷, et celle du *Language Observatory Project*²⁸. Nous nous trouvons pourtant face à divers facteurs nouveaux, notamment la taille extrême de la Toile et les dernières évolutions des moteurs de recherche qui ont perdu leur sérieux documentaire et la possibilité de couvrir une proportion importante des contenus, qui ne permettent plus d'obtenir des statistiques fiables et universelles. Les usages sociaux (Twitter, Facebook, Messagerie instantanée vocale et vidéo) et les sites multimédias échappent souvent aux analyses centrées sur le web. Cette situation de carence dans le domaine essentiel des indicateurs de la présence des langues dans le monde numérique a poussé le réseau Maaya à proposer la création d'un consortium

24 Dénomination de la génération utilisant Twitter, système ne permettant pas envoyer des messages de plus de 140 caractères.

25 Mathieu Olivier, « Facebook : l'arabe supplantera bientôt l'anglais », *Jeune Afrique*, 22 juillet 2011. <http://www.jeuneafrique.com/Article/ARTJAWEB20110722101339/algerie-liban-internet-facebookfacebook-l-arabe-supplantera-bientot-l-anglais.html>
Carrington Malin, *Facebook Arabic Rising*, 8 juin 2011. <http://www.spotonpr.com/facebook-arabic-uprising>

26 Voir dans cet ouvrage : Adel El Zaïm, *Cyberactivisme et langues régionales dans les printemps arabes de 2011*.

27 Voir l'étude en ligne http://dtil.unilat.org/LI/2007/fr/resultados_fr.htm

28 Voir dans cet ouvrage : Mikami Yoshiki & Shigeaki Kodama, *Mesurer la diversité linguistique sur le web*.

international dans le but de mener un projet ambitieux de cybermétrique des langues, un projet baptisé Dilinet.

Il s'agit de développer un ensemble de méthodes pour produire de manière pérenne des indicateurs de diversité linguistique sur internet, envisagé dans toute sa diversité, afin de servir de base à des politiques publiques dans tous les domaines liés à la société de l'information, à un niveau national et global.

Ce projet est une approche exploratoire de recherche, tenant compte des méthodes de mesure existantes en y ajoutant des approches novatrices, comprenant des systèmes de mesure témoignant du comportement des usagers. Pour surmonter les limitations créées par la taille de la Toile, le projet a besoin d'utiliser des méthodes d'indexation optimale basées sur des approches mathématiques non séquentielles ou statistiques, tout en ouvrant de nouvelles avenues telles que des techniques d'identification vocale ou la caractérisation automatique des contenus.

Le projet Dilinet reste transversal, à la fois champ de recherche (avec des chercheurs provenant des secteurs publics et privés) et outil de définition de politiques publiques. Il part de la motivation d'un groupe d'organismes internationaux tels que l'Unesco, l'Union internationale des Télécommunications, l'Organisation Internationale de la Francophonie (OIF) et l'Union Latine, et s'appuie sur l'expérience de Funredes et du Language Observatory Project cités plus haut.

En fournissant des résultats vérifiés et validés pour un ensemble d'indicateurs de diversité linguistique dans le cyberspace, Dilinet permettra de percevoir les tendances et de jauger les résultats des politiques mises en œuvre.

La mesure effective de la diversité linguistique dans le monde numérique contribuera au changement de paradigme dans la vision de la fracture numérique, en substituant à la mesure matérielle (réseaux et terminaux) la perspective de l'accès au contenu lui-même. Indirectement, Dilinet ouvrira également de nouvelles perspectives prometteuses pour la production d'indicateurs d'impact sur la société de l'information, et créera des opportunités pour que les langues soient considérées comme des paramètres importants de l'économie numérique.

BIBLIOGRAPHIE

[PIMIENTA 2009] D. Pimienta, D. Prado, Á. Blanco, *Douze années de mesure de la diversité linguistique sur l'Internet: bilan et perspectives*, Unesco, 2009. <http://unesdoc.unesco.org/images/0018/001870/187016f.pdf>.

[UIT 2009] UIT, *Measuring the Information Society: the ICT Development Index*, ISBN 92-61-12831-9, 2009.

[PAOLILLO, PIMIENTA, PRADO ET AL. 2005] J. Paolillo, D. Pimienta, D. Prado, et al. (2005), *Mesurer la diversité linguistique sur Internet*, Unesco, 12/2005. <http://unesdoc.unesco.org/images/0014/001421/142186f.pdf>.

[SUZUKI, MIKAMI ET AL. 2002] Suzuki I., Mikami Y., et al. (2002), "A Language and Character Set Determination Method Based on N-gram Statistics", dans *ACM Transactions on Asian Language Information Processing*, Vol.1, n°3.

[PRADO 2010] Prado, D. (2010) « Languages and Cyberspace: Analysis of the General Context and the Importance of Multilingualism in Cyberspace » In : *Proceedings of the International Conference Linguistic and Cultural Diversity in Cyberspace*, (Yakutsk, Russian Federation, 2-4 July, 2008). http://www.ifapcom.ru/files/Documents/multiling_eng.pdf

**L'AVENIR
SE PARLE,
S'ÉCRIT
ET SE LIT
DANS TOUTES
LES LANGUES**

CONCLUSION



ADAMA SAMASSÉKOU, président de Maaya.

ADAMA SAMASSÉKOU
PRÉSIDENT DE MAAVA

L'AVENIR SE
PARLE, S'ÉCRIT
ET SE LIT DANS
TOUTES LES
LANGUES

C'est par la langue que nous exprimons nos pensées et que nous communiquons avec autrui. Elle nous permet de partager notre expérience culturelle. Notre répertoire linguistique et notre choix des expressions déterminent qui nous sommes, le lieu et les circonstances du moment. Les langues sont ainsi les expressions vivantes des identités culturelles, individuelles et collectives.

Amadou Hampâté Ba disait que, de tous les éléments qui caractérisent l'individu, du physique au vestimentaire, c'est la langue qui demeure l'élément le plus évident, le plus pertinent.

Parce qu'elle transcende l'individu en faveur de sa communauté, la langue est par conséquent notre propriété aussi bien que celle de notre culture. C'est bien par elle que nous acquérons et transmettons notre savoir et notre savoir-faire qui permettent d'avoir une certaine emprise sur notre milieu. Elle constitue à la fois l'élément essentiel et le baromètre de notre développement : la langue est le constituant identitaire le plus fondamental.

LES LANGUES PORTENT EN ELLES LES SOURCES DE LA CRÉATIVITÉ

Socle de la culture et matrice de la créativité, la langue est l'instrument privilégié de construction des savoirs et de la connaissance. De ce point de vue, il est indéniable que la langue est l'une des meilleures expressions du génie créateur d'un peuple. La langue est le réceptacle et le véhicule de la culture. Elle est le vecteur, par excellence, de la cosmovision des sociétés humaines.

Évoquant cette question essentielle, le Professeur Joseph Ki-Zerbo, dans son ouvrage *À Quand l'Afrique?*, souligne : « *Le problème des langues est fondamental parce qu'il touche à l'identité des peuples. Et l'identité est nécessaire pour le développement comme pour la démocratie. Les langues touchent aussi à la culture, aux problèmes de la nation, à la capacité d'imaginer, à la créativité. Quand on répète dans une langue qui n'est pas originellement la sienne, on a une expression mécanique et mimétique de soi, sauf exceptions. (Mais gouverne-t-on pour les exceptions?) On ne fait qu'imiter. Alors que quand on s'exprime dans sa langue maternelle, l'imagination est libérée* »¹.

De son côté, Raymond Renard, auteur de l'ouvrage *Une éthique pour la francophonie. Questions de politique linguistique*, insiste sur l'importance de la sauvegarde de la diversité linguistique : « *Il est une raison, aussi fondamentale que celle du fondement identitaire, pour vouloir sauvegarder la diversité linguistique : c'est à travers leur propre langue que les individus voient le monde et qu'ils en interprètent le sens à leur manière, leur assurant ainsi l'accès à l'universel. Toutes les langues participent à l'interprétation de l'universel, car chaque culture produit des significations de valeur universelle. Et comme l'ont très bien démontré Alain Touraine ou Stephen Wurm, aucune langue, aucune culture ne pouvant prétendre représenter l'universel, mais chacune lui apportant sa propre contribution, c'est par le dialogue des langues et des cultures qu'on s'en rapprochera* »².

DIVERSITÉ LINGUISTIQUE ET CULTURELLE

Dialogue des langues et des cultures, c'est bien ce que prône l'Unesco, dans la « *Déclaration universelle de l'Unesco sur la diversité culturelle* »³ – texte fondateur s'il en est – qui considère la culture comme l'ensemble des traits distinctifs spirituels et matériels, intellectuels et affectifs qui caractérisent une société ou un groupe social donné. Ainsi définie, la culture reste au cœur de tous les débats sur l'identité, la cohésion sociale et le développement de toute économie fondée sur le savoir et le savoir-faire. C'est bien pour cette raison qu'elle constitue un patrimoine commun de l'Humanité.

1 Joseph Ki-Zerbo, *À Quand l'Afrique?* Éditions de l'Aube, 2003, pp. 81-82.

2 Raymond RENARD, *Une éthique pour la francophonie. Questions de politique linguistique*, Paris, Didier Érudition, Édition du CIPA, 2006.

3 Adoptée par la 31^e session de la Conférence générale de l'Unesco le 2 novembre 2001. <http://unesdoc.unesco.org/images/0012/001271/127160m.pdf>

Parce que chaque société, chaque groupe social est appelé à vivre sa culture, à préserver autant que possible ses valeurs culturelles séculaires, la diversité culturelle et linguistique devient donc une réalité qui s'impose à tous, au plan individuel, communautaire, national et mondial. C'est bien l'acceptation de cette réalité qui donne à chaque individu, à chaque communauté, à chaque société des droits humains fondamentaux, comme le stipule la Déclaration de l'Unesco, notamment :

- le droit de pouvoir s'exprimer, créer, et diffuser ses œuvres dans la langue de son choix, en particulier dans sa langue maternelle ;
- le droit à une éducation et une formation de qualité qui respectent pleinement son identité culturelle ;
- le droit de participer à la vie culturelle de son choix ;
- le droit d'exercer ses propres pratiques culturelles dans les limites qu'impose le respect des droits humains et des libertés fondamentales. Préserver et promouvoir la diversité et le pluralisme culturels et linguistiques au sein de tous les espaces demeure, de ce point de vue, une impérieuse nécessité.

L'enracinement de chaque groupe social ou linguistique dans ses propres valeurs culturelles, loin de constituer une source de division, contribue plutôt à renforcer la mobilisation des forces sociales. Comme ne cessent de le dire les Anciens, il n'y a pas meilleure connaissance que la connaissance de soi-même. En d'autres termes, pour mieux comprendre l'autre il faut d'abord se connaître soi-même. La compréhension mutuelle, d'abord au sein d'une même société, conduit à la cohésion sociale. Cet apprentissage acquis dans son milieu d'origine permet également une construction universelle.

En Afrique, cette attitude encore commune de coexistence est due à la conservation de principes éducatifs et de valeurs sociétales en vue de promouvoir la compréhension, le respect et la considération mutuels entre tous les individus. Toutes ces valeurs propres à la communauté, c'est-à-dire, la solidarité, le partage, la mesure, le consensus, l'entraide, l'accueil de l'Autre et l'hospitalité permettent l'acquisition de comportements et de notions mieux à même de prévenir les conflits et de les gérer.

La problématique de la diversité culturelle et linguistique, choix philosophico-politique bien ancré dans la vision africaine du monde, est bien résumée par l'écrivain et philosophe malien et africain, Amadou Hampâté Ba, dans cette citation : « *La beauté d'un tapis réside dans la variété de ses*

couleurs. S'il n'y a que du blanc, ce serait un drap blanc; s'il n'y a que du noir, ce serait un pagne de deuil. C'est l'Univers tout entier qui est notre patrie. Chacun de nous est une page du grand livre de la Nature. Dans la vaste communauté humaine lancée à la recherche d'un nouvel équilibre, chaque peuple doit apporter la note de son génie propre afin que tout l'ensemble en soit enrichi. Chacun doit s'ouvrir aux autres tout en restant lui-même».

De son côté, David Crystal écrit : « *La diversité occupe une place centrale dans les théories de l'évolution, car elle permet à une espèce de survivre dans des milieux différents. L'uniformisation présente des dangers pour la survie à long terme d'une espèce. Les écosystèmes les plus forts sont ceux qui sont les plus diversifiés. Si la multiplicité des cultures est une condition nécessaire pour un développement humain réussi, alors la préservation de la diversité linguistique est essentielle puisque les langues écrites et orales sont le principal mode de transmission des cultures* »⁴.

La pratique du multilinguisme constitue donc un des meilleurs moyens de préserver cette diversité.

MAAYA, HUMANITUDE

C'est pourquoi, il est important de s'engager résolument dans la valorisation du patrimoine culturel et linguistique des différents États par l'inventaire et la promotion de leurs langues, de leurs acquis culturels et artistiques, en vue de créer un cadre fécond de recherches et d'échanges sur les diverses valeurs culturelles et linguistiques. Alors, serons-nous en mesure d'enraciner les processus de développement dans la diversité et dans les forces endogènes de nos diverses cultures et sociétés.

Par ailleurs, faire le choix de préserver la diversité linguistique, c'est accepter de substituer à la logique marchande de concurrence destructrice une logique de solidarité et de complémentarité, susceptible de restaurer l'harmonie des êtres et des espèces. On le sait, la diversité culturelle et linguistique est à la société humaine ce que la biodiversité est à la nature : le ferment qui en garantit la vitalité, le socle de ce que nous avons appelé, il y a quelques années, notre « *humanité* », notre ouverture permanente à l'Autre, notre relation d'être humain à être humain, qui exige une relation solidaire permanente,

⁴ David Crystal, *À long terme, l'uniformisation menace l'espèce humaine*, Extrait du *Courrier International* n°486 du 24/02 au 01/03/2000.

sans calcul, un élan spontané d'accueil de l'Autre. Cette humanité permet de « *relier l'homme à l'homme* », selon la belle expression d'Aimé Césaire !

C'est par ce concept d'humanité que je traduis ce que nous appelons en Afrique *maaya* (en langue mandenkan), *neddaaku* (en fulfulde), *boroterey* (en songoy), *nite* (en wolof), *ubuntu* (dans les langues bantu)... Au Mandé on dit bien : « *Je suis un être humain non pas parce que je le pense, mais c'est ton regard posé sur moi qui fait de moi un être humain !* ».

Ainsi, promouvoir le multilinguisme c'est préserver la vitalité des sociétés humaines, c'est développer l'unicité de la communication humaine, c'est renforcer le dialogue entre les cultures, entre les peuples, c'est développer l'ouverture à l'Autre et lutter contre la violence et les crispations identitaires.

RÉUSSIR LE CYBERESPACE MULTILINGUE

Concernant les enjeux du multilinguisme dans le cyberspace et la problématique de la gouvernance globale, l'approche des langues que nous avons développée ici peut nous servir de guide et d'objectif.

Chaque langue reflétant la vision du monde de ses locuteurs, la diversité des langues du monde reflétée dans un internet multilingue peut préserver l'Humanité des dérives mécaniques d'une culture numérique désincarnée.

La promotion du multilinguisme dans le cyberspace pourrait contribuer à freiner le développement de « cellules individuelles de communication », de groupes isolés, et à renforcer le lien social, que le modèle dominant de développement peut affaiblir dangereusement.

La multiplicité des langues et l'aptitude à pratiquer plusieurs langues sont un puissant facteur de rapprochement des individus et des communautés. C'est pourquoi, garantir cette multiplicité des langues, à travers le cyberspace, dans la nouvelle société de la connaissance et des savoirs partagés en construction, facilitera l'acquisition d'une culture multilingue et d'une culture numérique démocratisée à chaque citoyen du monde.

Le multilinguisme étant dans le domaine de la culture ce que le multilatéralisme représente au plan politique, un des enjeux essentiels de cette culture multilingue est de contribuer à l'émergence d'une nouvelle gouvernance partenariale mondiale conduisant à la préservation de la Paix à l'échelle planétaire et à la sauvegarde de l'Humanité.

Conception graphique et mise en page : Nicolas Taffin et Kathleen Ponsard.

Ouvrage composé en *Minion* de Robert Slimbach et *Franklin Gothic* de Morris Fuller Benton. Deux classiques réunis qui nous vont comme une « petite robe noire » de typographie, pour le script dit latin.

L'arabe et le russe sont en *Times New Roman*. L'écriture devanagari est composée en *Devanagari Sangam*, le myanmar et le thaï en *Zawgyi-One*, fonte Unicode développée et maintenue en ligne par Alpha Computer en Copyleft (<http://code.google.com/p/zawgyi/wiki/MyanmarFontDownload>).

L'écriture kannada utilise *Kedage*, fonte Unicode adaptée par Nicholas Shanks à partir d'une police développée par l'Indian Institute of Science de Bangalore (<http://web.nickshanks.com/fonts/kannada/>). Cette police libre est encore inachevée, et sollicite les contributions en ligne des typographes de bonne volonté.

Couvrir l'ensemble des tables du projet Unicode avec des glyphes de qualité est une tâche considérable, mais offrir de surcroît un choix esthétique et une variété de styles aux utilisateurs relève du défi collectif. Le dessin typographique est un travail de grande précision dont dépend in fine la lisibilité des textes, dans toutes les cultures et pour toutes les écritures.

Dépôt légal 1^{er} trimestre 2012.
ISBN édition PDF 978-2-915825-23-7
<http://cfeditions.com>



NET.LANG EST DISPONIBLE EN PLUSIEURS LANGUES ET SUR PLUSIEURS SUPPORTS :

En version imprimée (français ou anglais)

480 pages, format 17 × 22,5 cm

prix : 34 euros

en français : ISBN 978-2-915825-08-4

en anglais : ISBN 978-2-915825-09-1

En librairie ou sur <http://cfeditions.com>

Au format ePub (français ou anglais)

sans DRM

en français : ISBN 978-2-915825-25-1

en anglais : ISBN 978-2-915825-26-8

<http://net-lang.net>

En pdf (français ou anglais)

sans DRM

en français : ISBN 978-2-915825-23-7

en anglais : ISBN 978-2-915825-24-4

<http://net-lang.net>

**D'AUTRES VERSIONS ET TRADUCTIONS
[HTTP://NET-LANG.NET](http://NET-LANG.NET)**

NET.LANG RÉUSSIR LE CYBERESPACE MULTILINGUE est un guide pédagogique, politique et pratique permettant d'appréhender et de comprendre les principaux enjeux du multilinguisme dans le cyberspace. Le multilinguisme est la nouvelle frontière du réseau numérique. Ce livre veut en montrer les enjeux, mais aussi proposer des pistes pour une présence équitable des langues dans la société de l'information.

Promouvoir sa langue, décider de politiques linguistiques, enseigner une ou des langues, travailler sur une langue, traduire ou interpréter des langues, créer du contenu dans plusieurs langues, s'adresser à des personnes d'autres langues... autant de situations pour lesquelles ce livre veut donner des pistes. La vitalité du multilinguisme est une force tant pour le développement de l'internet que pour la construction de sociétés inclusives, partageant les savoirs et les travaillant dans l'objectif du bien vivre ensemble.

Maaya / Réseau mondial pour la diversité linguistique est un réseau multi-acteurs créé dans le but de contribuer à la valorisation et la promotion de la diversité linguistique dans le monde. Terme de la langue bambara, Maaya pourrait se traduire par le néologisme « *Humanitude* ». Le Réseau Maaya a été créé dans le cadre du *Sommet Mondial sur la Société de l'Information* (SMSI), durant lequel la diversité culturelle et linguistique dans le cyberspace a été identifiée comme l'une des priorités. Maaya a été initiée par l'*Académie Africaine des Langues* (ACALAN), sous les auspices de l'*Union Africaine*.
<http://www.maaya.org/>

Ce livre est une réalisation du Réseau Maaya, avec le soutien de l'Unesco, Secteur de la Communication et de l'Information, de l'Organisation internationale de la Francophonie, de l'Union Latine, de ANLoc et de IDRC/CRDI.



ISBN (ÉDITION FRANÇAISE PDF) : 978-2-915825-23-7
<http://cfeditions.com>