

# Le web, une encyclopédie multilingue

*Marie Lebert, 2012*

*Diffusé par le Projet Gutenberg*

## Table des matières

- 1974 > Les débuts de l'internet
- 1986 > Des extensions pour l'ASCII
- 1990 > Le web booste l'internet
- 1990 > La LINGUIST List
- 1991 > L'Unicode, système d'encodage universel
- 1994 > Travlang, des langues pour voyager
- 1995 > L'Internet Dictionary Project
- 1995 > NetGlos, glossaire de l'internet
- 1995 > Plusieurs langues sur notre écran
- 1995 > Global Reach, pour localiser les sites web
- 1995 > OneLook Dictionaries, point d'accès commun
- 1997 > Un web anglophone à 82,3%
- 1997 > Une liste de langues européennes minoritaires
- 1997 > Une base terminologique européenne
- 1997 > Babel Fish, logiciel de traduction gratuit
- 1997 > Les outils de la société de traduction Logos
- 1997 > Des bases terminologiques spécialisées
- 1998 > La nécessité d'une «démocratie linguistique»
- 1999 > Les dictionnaires bilingues de WordReference.com
- 1999 > L'internet, outil indispensable pour les traducteurs
- 1999 > La nécessité d'une information bilingue
- 2000 > Encyclopédies et dictionnaires en ligne
- 2000 > Le portail yourDictionary.com
- 2000 > Le Projet Gutenberg et les langues
- 2001 > Wikipédia, encyclopédie collaborative
- 2001 > L'UNL, projet de métalangage numérique
- 2001 > Un marché pour les logiciels de traduction
- 2004 > Le web 2.0, communauté et partage
- 2007 > La norme ISO 639-3 pour identifier les langues
- 2007 > Google Traduction
- 2009 > 6.909 langues vivantes dans l'Ethnologue
- 2010 > Un atlas de l'UNESCO pour les langues menacées

## Introduction

«Le web sera une encyclopédie du monde faite par le monde pour le monde. Il n'y aura plus d'informations ni de connaissances utiles qui ne soient pas disponibles, si bien que l'obstacle principal à la compréhension internationale et interpersonnelle et au développement personnel et institutionnel sera levé. Il faudrait une imagination plus débordante que la mienne pour prédire l'effet de ce développement sur l'humanité.» (Robert Beard, créateur du site A Web of Online Dictionaries, septembre 1998)

Ce livre se présente sous la forme d'une chronologie en 32 chapitres de 1974 à 2010. Merci à toutes les personnes qui sont citées ici, pour leur temps et pour leur amitié. Sauf indication contraire, les citations proviennent des Entretiens du NEF (Net des études françaises), menés par l'auteure au fil des ans.

## 1974 > Les débuts de l'internet

### **[Résumé]**

L'internet naît en 1974, quinze ans avant le web. Vinton Cerf est souvent appelé le père de l'internet parce qu'il est le co-auteur en 1974 avec Bob Kahn du protocole TCP/IP (Transmission Control Protocol / Internet Protocol) nécessaire au bon fonctionnement du réseau. L'internet est d'abord mis en place aux États-Unis pour relier les agences gouvernementales, les universités et les centres de recherche, avant de débiter sa progression mondiale en 1983. L'internet trouve ensuite un nouveau souffle avec l'invention du web par Tim Berners-Lee en 1990 puis le lancement du premier navigateur Mosaic en 1993. Vinton Cerf fonde l'Internet Society (ISOC) en 1992 pour promouvoir le développement du réseau. Interviewé en janvier 1998 par le quotidien Libération, il explique: «Le réseau fait deux choses (...): comme les livres, il permet d'accumuler de la connaissance. Mais, surtout, il la présente sous une forme qui la met en relation avec d'autres informations. Alors que, dans un livre, l'information est maintenue isolée.»

\*\*\*

L'internet naît en 1974 suite à la création du protocole TCP/IP (Transmission Control Protocol / Internet Protocol) par Vinton Cerf et Bob Kahn pour les échanges de données sur le réseau, quinze ans avant l'invention du web.

### ***Les premiers pas***

Vinton Cerf est souvent appelé le père de l'internet parce qu'il est le co-auteur en 1974 avec Bob Kahn du protocole TCP/IP (Transmission Control Protocol / Internet Protocol) nécessaire au bon fonctionnement du réseau.

L'internet est d'abord mis en place aux États-Unis pour relier les agences gouvernementales, les universités et les centres de recherche, avant de débiter sa

progression mondiale en 1983. Il trouve ensuite un nouveau souffle avec l'invention du web par Tim Berners-Lee en 1990 puis le lancement du premier navigateur Mosaic en 1993.

Vinton Cerf fonde l'Internet Society (ISOC) en 1992 pour promouvoir le développement du réseau. Interviewé en janvier 1998 par le quotidien Libération, il explique: «Le réseau fait deux choses (...): comme les livres, il permet d'accumuler de la connaissance. Mais, surtout, il la présente sous une forme qui la met en relation avec d'autres informations. Alors que, dans un livre, l'information est maintenue isolée.»

Le web étant facile d'utilisation grâce aux liens hypertextes reliant les documents entre eux, l'internet peut enfin être utilisé par le grand public dans les années 1990, et pas seulement par les usagers versés dans l'informatique. On compte 100 millions d'utilisateurs en décembre 1997, avec un million de nouveaux utilisateurs par mois, et 300 millions d'utilisateurs en décembre 2000.

### ***La situation en Europe***

En ce qui concerne la connexion à l'internet, les choses sont moins faciles en Europe qu'en Amérique du Nord. La connexion est d'abord tarifée à la durée, avec un tarif de jour très élevé et un tarif de nuit plus intéressant, d'où l'obligation de travailler la nuit pour éviter les factures trop élevées. Des mouvements de grève sont lancés fin 1998 et début 1999 en France, en Italie et en Allemagne dans le but de faire pression sur les sociétés prestataires pour qu'elles baissent leurs prix et qu'elles proposent des forfaits internet, avec gain de cause les mois suivants.

Quelques années plus tard, le haut débit se généralise. Jean-Paul, webmestre du site hypermédia cotres.net, résume la situation en janvier 2007: «J'ai l'impression que nous vivons une période "flottante", entre les temps héroïques, où il s'agissait d'avancer en attendant que la technologie nous rattrape, et le futur, où le très haut débit va libérer les forces qui commencent à bouger, pour l'instant dans les seuls jeux.»

### ***L'internet du futur***

L'internet du futur pourrait être un réseau pervasif permettant de se connecter en tout lieu et à tout moment sur tout type d'appareil à travers un réseau unique et omniprésent.

Le concept de réseau pervasif est développé par Rafi Haladjian, fondateur de la société Ozone. Comme expliqué sur le site web en 2007, «la nouvelle vague touchera notre monde physique, notre environnement réel, notre vie quotidienne dans tous les instants. Nous n'accéderons plus au réseau, nous l'habiterons. Les composantes futures de ce réseau (parties filiales, parties non filiales, opérateurs) seront transparentes à l'utilisateur final. Il sera toujours ouvert, assurant une permanence de la connexion en tout lieu. Il sera également agnostique en terme d'application(s), puisque fondé sur les protocoles mêmes de l'internet.» Nous attendons cela avec impatience.

Quant au contenu de l'internet, Timothy Leary, philosophe visionnaire, le décrit ainsi dans son livre «Chaos et cyberculture?», publié en 1994: «Toute l'information du monde est à l'intérieur. Et grâce au cyberspace, tout le monde peut y avoir accès. Tous les signaux humains contenus jusque-là dans les livres ont été numérisés. Ils sont enregistrés et disponibles dans ces banques de données, sans compter tous les tableaux, tous les films, toutes les émissions de télé, tout, absolument tout.» En 2011, nous n'en sommes pas encore là, mais les choses sont en bonne voie.

## **1986 > Des extensions pour l'ASCII**

### **[Résumé]**

Avec le développement de l'internet hors de la sphère anglophone, communiquer uniquement en anglais devient insuffisant, d'où la nécessité de prendre en compte les caractères accentués d'autres langues européennes. Publié par l'American National Standards Institute (ANSI) en 1963, l'ASCII (American Standard Code for Information Interchange) est le premier système d'encodage. Il s'agit d'un code standard de 128 caractères traduits en langage binaire sur sept bits (A est traduit par «1000001», B est traduit par «1000010», etc.). L'ASCII permet uniquement la lecture de l'anglais (et du latin). Des variantes de l'ASCII sur huit bits sont publiées à partir de 1986 pour prendre en compte les caractères accentués de quelques langues européennes. La variante pour le français, l'espagnol et l'allemand (entre autres) est la norme ISO 8859-1 (Latin-1). Mais les problèmes sont loin d'être résolus. Pour cela, il faudra attendre l'Unicode, nouveau système d'encodage universel dont la première version est publiée en janvier 1991.

\*\*\*

Avec le développement de l'internet hors de la sphère anglophone, communiquer uniquement en anglais devient insuffisant, d'où la nécessité de prendre en compte les caractères accentués de plusieurs langues européennes.

### ***L'ASCII sur 7 bits***

Le premier système d'encodage informatique est l'ASCII (American Standard Code for Information Interchange). Publié en 1963 aux États-Unis par l'American National Standards Institute (ANSI), l'ASCII est un code standard de 128 caractères traduits en langage binaire sur sept bits (A est traduit par «1000001», B est traduit par «1000010», etc.). Les 128 caractères comprennent 33 caractères de contrôle (qui ne représentent donc pas de symbole écrit) et 95 caractères imprimables: les 26 lettres sans accent en majuscules (A-Z) et minuscules (a-z), les chiffres, les signes de ponctuation et quelques caractères spéciaux, le tout correspondant aux touches du clavier anglophone.

### ***L'ASCII sur 8 bits***

L'ASCII permet uniquement la lecture de l'anglais (et du latin). L'ASCII ne permet donc pas de prendre en compte les lettres accentuées présentes dans bon nombre

de langues européennes (français, espagnol, allemand, etc.), tout comme les langues disposant d'autres alphabets (arabe, grec, russe, etc.) et à plus forte raison les langues non alphabétiques (chinois, coréen, japonais, etc.). Ceci ne pose pas de problème majeur les premières années, tant que l'échange de fichiers électroniques se limite surtout à l'Amérique du Nord. Mais le multilinguisme devient bientôt une nécessité vitale. Des variantes de l'ASCII sur huit bits sont publiées à partir de 1986 pour prendre en compte les caractères accentués de quelques langues européennes. La variante pour le français, l'espagnol et l'allemand (entre autres) est la norme ISO 8859-1 (ISO Latin-1).

### ***Un véritable casse-tête***

Avec le développement de l'internet, l'échange des données s'internationalise encore davantage. Même avec des variantes de l'ASCII, on ne peut décidément plus se limiter à l'utilisation d'un système d'encodage datant des débuts de l'informatique. De plus, le passage de l'ASCII original à ses différentes variantes devient vite un véritable casse-tête, y compris au sein de l'Union européenne, les problèmes étant entre autres la multiplication des variantes, la corruption des données dans les échanges informatiques ou encore l'incompatibilité des systèmes, les pages ne pouvant être affichées que dans une seule langue à la fois.

Olivier Gainon, fondateur de CyLibris et pionnier de l'édition électronique littéraire, écrit à ce sujet en décembre 2000: «Il faut que le réseau respecte les lettres accentuées, les lettres spécifiques, etc. Je crois très important que les futurs protocoles permettent une transmission parfaite de ces aspects - ce qui n'est pas forcément simple (dans les futures évolutions de l'HTML ou des protocoles IP, etc.). Donc il faut que chacun puisse se sentir à l'aise avec l'internet et que ce ne soit pas simplement réservé à des (plus ou moins) anglophones. Il est anormal aujourd'hui que la transmission d'accents puisse poser problème dans les courriers électroniques. La première démarche me semble donc une démarche technique. Si on arrive à faire cela, le reste en découle: la représentation des langues se fera en fonction du nombre de connectés, et il faudra envisager à terme des moteurs de recherche multilingues.»

### ***L'Unicode***

Publié pour la première fois en janvier 1991, l'Unicode est un système d'encodage universel sur 16 bits spécifiant un nombre unique pour chaque caractère. Ce nombre est lisible quels que soient la plateforme, le logiciel et la langue utilisés. L'Unicode peut traiter 65.000 caractères uniques et prendre en compte tous les systèmes d'écriture de la planète. L'Unicode est progressivement adopté à partir de 1998. Un énorme travail est en effet nécessaire pour sa prise en compte par tous les logiciels et navigateurs web. Il faudra attendre décembre 2007 pour que l'Unicode supplante l'ASCII sur l'internet.

## **1990 > Le web booste l'internet**

### ***[Résumé]***

Le World Wide Web est inventé en 1990 par Tim Berners-Lee, alors chercheur au CERN (Centre européen pour la recherche nucléaire) à Genève, en Suisse. En 1989, il met au point l'hypertexte pour relier des documents entre eux. En 1990, il met au point le premier serveur HTTP (HyperText Transfer Protocol) et le premier navigateur web. En 1991, le web est opérationnel et rend l'internet (qui existe depuis 1974) accessible à tous et pas seulement aux usagers versés dans l'informatique. Des liens hypertextes permettent désormais de passer d'un document textuel ou visuel à un autre au moyen d'un simple clic de souris. Plus tard, cette interactivité est encore accrue avec la possibilité de liens hypermédias permettant de lier des textes et des images à des vidéos ou bandes sonores. Le World Wide Web Consortium (W3C) est fondé en octobre 1994 pour développer les protocoles communs du web.

\*\*\*

Le World Wide Web est inventé en 1990 par Tim Berners-Lee, chercheur au CERN (Centre européen pour la recherche nucléaire) à Genève, en Suisse. Le web rend l'internet accessible à tous et lui permet une progression exponentielle.

### ***Les débuts du web***

En 1989, Tim Berners-Lee met au point l'hypertexte pour relier des documents entre eux. En 1990, il met au point le premier serveur HTTP (HyperText Transfer Protocol) et le premier navigateur web. En 1991, le World Wide Web est opérationnel et rend l'internet (qui existe depuis 1974) accessible à tous et pas seulement aux usagers versés dans l'informatique. Des liens hypertextes permettent désormais de passer d'un document textuel à un autre au moyen d'un clic de souris. Plus tard, cette interactivité est encore accrue avec la possibilité de liens hypermédias permettant de lier textes et images fixes à des vidéos ou bandes sonores.

Mosaic est le premier navigateur destiné au grand public. Développé par le NSCA (National Center for Supercomputing Applications) à l'Université de l'Illinois (États-Unis) et distribué gratuitement en novembre 1993, il contribue largement au développement rapide du web. Début 1994, une partie de l'équipe de Mosaic émigre dans la Netscape Communications Corporation pour développer un nouveau logiciel sous le nom de Netscape Navigator. En 1995, Microsoft lance son propre navigateur, l'Internet Explorer. Viennent ensuite d'autres navigateurs, comme Opera ou Safari, le navigateur d'Apple.

Un consortium industriel international est fondé en octobre 1994 pour développer les protocoles communs du web, sous le nom de World Wide Consortium (W3C) et sous l'égide de Tim Berners-Lee. En 1997, une section Internationalization / Localization regroupe les protocoles utilisés pour créer un site web multilingue: HTML (HyperText Markup Language), jeux (de base) de caractères, nouveaux attributs, HTTP (HyperText Transfer Protocol), négociation de la langue, URL (Uniform Resource Locator) et autres identificateurs incluant des caractères non ASCII, conseils divers.

## **Le rêve de Tim Berners-Lee**

À la question de Pierre Ruetschi, journaliste à la Tribune de Genève, quotidien suisse: «Sept ans plus tard, êtes-vous satisfait de la façon dont le web a évolué?», Tim Berners-Lee répond en décembre 1997 que, s'il est heureux de la richesse et de la variété de l'information disponible, le web n'a pas encore la puissance prévue dans sa conception d'origine. Il aimerait «que le web soit plus interactif, que les gens puissent créer de l'information ensemble», et pas seulement consommer celle qui leur est proposée. Le web doit devenir «un média de collaboration, un monde de connaissance que nous partageons».

Dans un essai publié en avril 1998 sur sa propre page web (sur le site du World Wide Web Consortium), Tim Berners-Lee explique que «le rêve derrière le web est un espace d'information commun dans lequel nous communiquons en partageant l'information. Son universalité est essentielle, à savoir le fait qu'un lien hypertexte puisse pointer sur quoi que ce soit, quelque chose de personnel, de local ou de global, aussi bien une ébauche qu'une réalisation très sophistiquée. Deuxième partie de ce rêve, le web deviendrait d'une utilisation tellement courante qu'il serait un miroir réaliste (sinon la principale incarnation) de la manière dont nous travaillons, jouons et nouons des relations sociales. Une fois que ces interactions seraient en ligne, nous pourrions utiliser nos ordinateurs pour nous aider à les analyser, donner un sens à ce que nous faisons, et voir comment chacun trouve sa place et comment nous pouvons mieux travailler ensemble.» (extrait de «The World Wide Web: a very short personal history»)

## **Le web 2.0**

Selon Netcraft, société spécialisée dans les mesures d'audience, le nombre de sites web passe d'un million de sites (avril 1997) à dix millions de sites (février 2000), 20 millions de sites (septembre 2000), 30 millions de sites (juillet 2001), 40 millions de sites (avril 2003), 50 millions de sites (mai 2004), 60 millions de sites (mars 2005), 70 millions de sites (août 2005), 80 millions de sites (avril 2006), 90 millions de sites (août 2006) et 100 millions de sites (novembre 2006), une augmentation rapide favorisée par l'explosion des sites personnels et des blogs.

Le web 2.0, terme lancé en 2004 par Tim O'Reilly, éditeur de livres informatiques, apporte peut-être un début de réponse au rêve de Tim Berners-Lee puisqu'il est basé sur les notions de communauté et de partage.

Quinze ans après la création du web, le magazine Wired constate dans son numéro d'août 2005 que «moins de la moitié du web est commercial, le reste fonctionne avec la passion». Quant à l'internet, d'après le quotidien Le Monde du 19 août 2005, «ses trois pouvoirs - l'ubiquité, la variété et l'interactivité - rendent son potentiel d'usages quasi infini».

Robert Beard, professeur de langues et créateur du site A Web of Online Dictionaries en 1995, écrivait de manière prémonitoire dès septembre 1998: «Le web sera une encyclopédie du monde faite par le monde pour le monde. Il n'y aura plus d'informations ni de connaissances utiles qui ne soient pas disponibles, si bien que l'obstacle principal à la compréhension internationale et interpersonnelle et au

développement personnel et institutionnel sera levé. Il faudrait une imagination plus débordante que la mienne pour prédire l'effet de ce développement sur l'humanité.»

## 1990 > La LINGUIST List

### **[Résumé]**

Liste de diffusion à destination des linguistes, la LINGUIST List est créée par Anthony Rodrigues Aristar en 1990 au sein de l'University of Western Australia. Avec 60 inscrits, la liste déménage vers la Texas A&M University (États-Unis) en 1991, l'Eastern Michigan University étant le principal éditeur. En 1997, la LINGUIST List dispose de son propre site web. Les messages reçus sont classés dans diverses rubriques: profession (conférences, associations linguistiques, programmes), recherche et soutien à la recherche (articles, résumés de mémoires, projets, bibliographies, sujets, textes), publications, pédagogie, ressources linguistiques (langues, familles linguistiques, dictionnaires, informations régionales) et soutien informatique (polices de caractères et logiciels). La LINGUIST List est une composante de la WWW Virtual Library pour la linguistique.

\*\*\*

Liste de diffusion à destination des linguistes, la LINGUIST List est créée par Anthony Rodrigues Aristar en 1990 au sein de l'University of Western Australia.

Avec 60 inscrits, la liste déménage vers la Texas A&M University (États-Unis) en 1991, l'Eastern Michigan University étant le principal éditeur.

En 1997, la LINGUIST List dispose de son propre site web. Les messages reçus sont classés dans diverses rubriques: profession (conférences, associations linguistiques, programmes), recherche et soutien à la recherche (articles, résumés de mémoires, projets, bibliographies, sujets, textes), publications, pédagogie, ressources linguistiques (langues, familles linguistiques, dictionnaires, informations régionales) et soutien informatique (polices de caractères et logiciels). La LINGUIST List est une composante de la WWW Virtual Library pour la linguistique.

Helen Dry, co-modératrice de la LINGUIST List depuis 1991, explique en août 1998: «La LINGUIST List, que je modère, a pour politique d'accepter les informations dans toutes les langues, puisque c'est une liste pour linguistes. Nous ne souhaitons cependant pas que le message soit publié dans plusieurs langues, tout simplement à cause de la charge de travail que cela représenterait pour notre personnel de rédaction. (Nous ne sommes pas une liste fourre-tout, mais une liste modérée. De ce fait, avant d'être publié, chaque message est classé par nos étudiants-rédacteurs dans une section comprenant des messages du même type.) Notre expérience nous montre que pratiquement tout le monde choisit de publier en anglais. Mais nous relient ces informations à un système de traduction présentant nos pages dans cinq langues différentes. Ainsi un abonné ne lit LINGUIST en anglais que s'il le souhaite. Nous essayons aussi d'avoir au moins un étudiant-éditeur qui soit réellement multilingue, afin que les lecteurs puissent correspondre avec nous dans d'autres langues que l'anglais.»

Helen Dry ajoute en juillet 1999: «Nous commençons maintenant à rassembler un grand nombre de données. Par exemple, nous avons des bases de données avec moteur de recherche pour les résumés de thèses de linguistique, pour les informations sur les programmes universitaires de linguistique et pour les données professionnelles de linguistes individuels. À ma connaissance, le fichier des résumés de thèses est la seule compilation électronique qui soit disponible gratuitement sur l'internet.»

## **1991 > L'Unicode, système d'encodage universel**

### **[Résumé]**

L'ASCII, premier système d'encodage datant des débuts de l'informatique, n'est plus suffisant avec l'internationalisation de l'internet, d'où l'intérêt de l'Unicode, nouveau système d'encodage universel, dont la première version est publiée en janvier 1991. L'Unicode spécifie un nombre sur 16 bits unique à chaque caractère (ou idéogramme) et lisible quels que soient la plateforme, le logiciel et la langue utilisés. L'Unicode peut traiter 65.000 caractères et prendre en compte tous les systèmes d'écriture de la planète. Il devient une composante des spécifications du World Wide Web Consortium (W3C), l'organisme international chargé du développement du web. L'utilisation de l'Unicode se généralise à partir de 1998, par exemple pour les fichiers texte sous plateforme Windows (Windows NT, Windows 2000, Windows XP et versions suivantes), qui étaient jusque-là en ASCII. L'Unicode supplante définitivement l'ASCII en décembre 2007.

\*\*\*

L'ASCII n'est plus suffisant avec l'internationalisation de l'internet, d'où l'intérêt de l'Unicode, nouveau système d'encodage universel, dont la première version est publiée en janvier 1991.

Contrairement à l'ASCII conçu pour l'anglais (et le latin), avec des variantes pour quelques langues supplémentaires, l'Unicode prend en compte toutes les langues de la planète.

### **De l'ASCII à l'Unicode**

Pour mémoire, le premier système d'encodage informatique est l'ASCII (American Standard Code for Information Interchange), publié en 1963 aux États-Unis par l'American National Standards Institute (ANSI) pour encoder des informations en anglais.

Mais le multilinguisme devient bientôt une nécessité vitale. Des variantes de l'ASCII prennent en compte d'autres langues à partir de 1986. Avec le développement de l'internet, l'échange des données s'internationalise de plus en plus, si bien qu'il n'est plus possible de se limiter à un système d'encodage datant des débuts de l'informatique, même avec ses variantes.

Publié pour la première fois en janvier 1991, l'Unicode est un système d'encodage universel sur 16 bits spécifiant un nombre unique pour chaque caractère (ou idéogramme). Ce nombre est lisible quels que soient la plateforme, le logiciel et la langue utilisés. L'Unicode peut traiter 65.000 caractères uniques et prendre en compte tous les systèmes d'écriture de la planète. À la grande satisfaction des linguistes, il remplace progressivement l'ASCII, avec des variantes UTF-8, UTF-16 et UTF-32 (UTF: Unicode Transformation Format) selon le nombre de bits utilisés pour l'encodage.

L'Unicode est maintenu par l'Unicode Consortium. Il devient une composante des spécifications du World Wide Web Consortium (W3C), fondé en octobre 1994 pour promouvoir le développement du web. L'utilisation de l'Unicode se généralise à partir de 1998, par exemple pour les fichiers texte sous plateforme Windows (Windows NT, Windows 2000, Windows XP et versions suivantes), qui étaient jusque-là en ASCII.

### ***Une tâche énorme***

Mais la tâche s'annonce rude. Patrick Rebollar, professeur de français et de littérature française au Japon et modérateur de la liste de diffusion LITOR (Littérature et ordinateur), précise en janvier 2000: «Il s'agit d'abord d'un problème logiciel. Comme on le voit avec Netscape ou Internet Explorer, la possibilité d'affichage multilingue existe. La compatibilité entre ces logiciels et les autres (de la suite Office de Microsoft, par exemple) n'est cependant pas acquise. L'adoption de la table Unicode devrait résoudre une grande partie des problèmes, mais il faut pour cela réécrire la plupart des logiciels, ce à quoi les producteurs de logiciels rechignent du fait de la dépense, pour une rentabilité qui n'est pas évidente car ces logiciels entièrement multilingues intéressent moins de clients que les logiciels de navigation.»

Luc Dall'Armellina, co-auteur et webmestre d'oVosite, un espace d'écriture hypermédia, écrit en juin 2000: «Les systèmes d'exploitation se dotent peu à peu des kits de langues et bientôt peut-être de polices de caractères Unicode à même de représenter toutes les langues du monde; reste que chaque application, du traitement de texte au navigateur web, emboîte ce pas. Les difficultés sont immenses: notre clavier avec ses ± 250 touches avoue ses manques dès lors qu'il faille saisir des Katakana ou Hiragana japonais, pire encore avec la langue chinoise. La grande variété des systèmes d'écriture de par le monde et le nombre de leurs signes font barrage. Mais les écueils culturels ne sont pas moins importants, liés aux codes et modalités de représentation propres à chaque culture ou ethnie.» Un sentiment prémonitoire puisque l'Unicode ne supplantera l'ASCII qu'en décembre 2007.

## **1994 > Travlang, des langues pour voyager**

### ***[Résumé]***

Les premiers dictionnaires de langues en ligne sont d'un niveau très moyen, mais il faut un début à tout et ils dépannent les voyageurs. En 1994, Michael C. Martin, étudiant en physique, crée d'abord une rubrique intitulée «Foreign Languages

for Travelers» sur le site de son université à New York. L'année suivante, il lance Travlang, un site dédié à la fois aux voyages et aux langues, nommé meilleur site de voyages en 1997. En août 1998, la section «Foreign Languages for Travelers» permet d'apprendre les rudiments de soixante langues sur le web. La section «Translating Dictionaries» donne accès à des dictionnaires gratuits dans quinze langues (afrikaans, allemand, danois, espagnol, espéranto, finnois, français, frison, hollandais, hongrois, italien, latin, norvégien, portugais, tchèque). Ces dictionnaires sont le plus souvent sommaires et de qualité inégale. D'autres sections offrent des liens vers des services de traduction, des écoles de langue, des librairies multilingues, etc.

\*\*\*

Les premiers dictionnaires de langues disponibles sur le web - tels ceux de Travlang - sont d'un niveau très moyen, mais il faut un début à tout. Ils dépannent les voyageurs et fraient une voie aux bons sites de dictionnaires qui suivront ensuite.

En 1994, Michael C. Martin, étudiant en physique, crée d'abord une rubrique intitulée «Foreign Languages for Travelers» sur le site de son université à New York, pour apprendre les langues sur le web. Cette rubrique s'étoffe rapidement et rencontre un grand succès. L'année suivante, il lance Travlang, un site dédié à la fois aux voyages et aux langues, nommé meilleur site de voyages en 1997. Devenu chercheur en physique au Lawrence Berkeley National Laboratory (Californie), Michael continue de gérer Travlang lui-même.

En 1998, la section «Foreign Languages for Travelers» permet d'apprendre les rudiments de soixante langues sur le web. La section «Translating Dictionaries» donne accès à des dictionnaires gratuits dans quinze langues (afrikaans, allemand, danois, espagnol, espéranto, finnois, français, frison, hollandais, hongrois, italien, latin, norvégien, portugais, tchèque). Ces dictionnaires sont le plus souvent sommaires et de qualité inégale.

D'autres sections offrent des liens vers des services de traduction, des écoles de langue, des librairies multilingues, etc. On peut également réserver son hôtel, sa voiture ou son billet d'avion, s'informer des taux de change ou encore consulter un index de 7.000 liens vers d'autres sites de langues et de voyages.

Michael C. Martin écrit en août 1998: «Je pense que le web est un endroit idéal pour rapprocher les cultures et les personnes, et ceci inclut d'être multilingue. Notre site Travlang est très populaire pour cette raison, et les gens aiment le contact avec d'autres parties du monde. (...) L'internet est vraiment un outil important pour communiquer avec des gens avec lesquels on n'aurait pas l'occasion de dialoguer autrement. J'apprécie vraiment la collaboration générale qui a rendu possibles les pages de Foreign Languages for Travelers.»

En ce qui concerne l'avenir, «je pense que les traductions intégrales informatisées vont devenir monnaie courante, et qu'elles permettront de communiquer à la base avec davantage de gens. Ceci aidera aussi à amener davantage l'internet au monde non anglophone.»

Travlang est acquis en février 1999 par la société GourmetMarket.com, puis racheté en janvier 2000 par la société iiGroup. En juillet 2000, le site compte 2 millions de visiteurs par mois. C'est l'époque où on commence à trouver de bons dictionnaires sur l'internet... mais sur d'autres sites, Travlang étant devenu commercial à 200%, dans le pire sens du terme, est désormais uniquement axé sur les voyages.

## 1995 > L'Internet Dictionary Project

### **[Résumé]**

Tyler Chambers, informaticien, crée d'abord en mai 1994 la Human-Languages Page (H-LP) pour proposer un index des ressources linguistiques disponibles sur l'internet dans diverses langues. Puis il lance en parallèle en 1995 l'Internet Dictionary Project (IDP), à savoir un projet coopératif ouvert à tous pour constituer des dictionnaires en accès libre sur le web, de l'anglais vers d'autres langues (allemand, espagnol, français, italien, latin, portugais). Comme expliqué sur le site web, «le but est de créer des dictionnaires de traduction grâce à l'aide des internautes. Ce site permet aux usagers du monde entier de les consulter et de participer à la traduction de termes anglais dans d'autres langues. Les listes de termes anglais et leurs correspondants dans d'autres langues sont ensuite mis à la disposition de tous sur ce site, sans restriction d'aucune sorte.»

\*\*\*

Tyler Chambers, informaticien, lance l'Internet Dictionary Project (IDP) en 1995 en tant que projet coopératif ouvert à tous pour constituer des dictionnaires en accès libre sur le web.

Un an auparavant, Tyler crée la Human-Languages Page (H-LP) pour proposer un répertoire de ressources linguistiques. Ce répertoire recense 1.800 ressources dans une centaine de langues en octobre 1998. Ces ressources sont classées dans plusieurs rubriques: langues et littérature, écoles et institutions, ressources linguistiques, produits et services, organismes, emplois et stages, dictionnaires, cours de langues.

Quant à l'Internet Dictionary Project, son but est de créer des dictionnaires de traduction de l'anglais vers d'autres langues (allemand, espagnol, français, italien, latin, portugais).

Comme expliqué à l'époque sur le site web, «le but est de créer des dictionnaires de traduction grâce à l'aide des internautes. Ce site permet aux usagers du monde entier de les consulter et de participer à la traduction de termes anglais dans d'autres langues. Les listes de termes anglais et leurs correspondants dans d'autres langues sont ensuite mis à la disposition de tous sur ce site, sans restriction d'aucune sorte. (...)

L'Internet Dictionary Project a débuté en 1995 pour combler une lacune et procurer des dictionnaires de traduction gratuits à la communauté des internautes et à tous ceux qui s'intéressent à l'informatique. Non seulement il est très utile d'avoir

immédiatement accès à des dictionnaires par le World Wide Web, mais ceci permet aussi le développement de logiciels pouvant tirer parti de tels dictionnaires, que ce soit des programmes de traduction ou des vérificateurs d'orthographe ou encore des guides d'apprentissage des langues. En facilitant la création de ces dictionnaires en ligne par des milliers de volontaires, et en les mettant gratuitement à la disposition de tous, l'Internet Dictionary Project espère imprimer sa marque sur l'internet et susciter d'autres projets qui seront plus bénéfiques que de générer des revenus purement financiers.»

Tyler Chambers raconte en septembre 1998 lors d'un entretien par courriel: «Le multilinguisme sur le web était inévitable bien avant que ce médium ne se développe vraiment. Mon premier vrai contact avec le web date de 1994, un peu après ses débuts mais bien avant son expansion. 1994 a été aussi l'année où j'ai débuté mon premier projet web multilingue, et il existait déjà un nombre significatif de ressources linguistiques en ligne. Ceci était antérieur à la création de Netscape. Mosaic était le seul navigateur sur le web, et les pages web étaient essentiellement des documents textuels reliés par des hyperliens. Avec l'amélioration des navigateurs et l'expérience acquise par les usagers, je ne pense pas qu'il existe une langue vivante qui ne soit pas maintenant représentée sur le web, que ce soit la langue des Indiens d'Amérique ou les dialectes moyen-orientaux. De même une pléthore de langues mortes peut maintenant trouver une audience nouvelle avec des érudits et autres spécialistes en ligne. (...)

Bien que je ne sois pas multilingue, ni même bilingue moi-même, je suis conscient du fait que très peu de domaines ont une importance comparable à celle des langues et du multilinguisme. (...) Dans l'ensemble, je pense que le web est important pour la sensibilisation aux langues et pour les questions culturelles. Dans quel autre endroit peut-on chercher au hasard pendant vingt minutes et trouver des informations susceptibles de vous intéresser dans trois langues différentes sinon plus? (...)

Dire que l'internet aiguillonne le multilinguisme est à mon sens une opinion fautive. C'est la communication qui aiguillonne le multilinguisme et l'échange multiculturel. L'internet est seulement le mode de communication le plus récent qui soit accessible aux gens plus ou moins ordinaires. (...) Les langues deviendront encore plus importantes qu'elles ne le sont lorsque tout le monde pourra communiquer à l'échelle de la planète (à travers le web, les discussions, les jeux, le courrier électronique, ou toute application appartenant encore au domaine de l'avenir).»

Au printemps 2001, la Human-Languages Page fusionne avec le Languages Catalog, une section de la WWW Virtual Library, pour devenir iLoveLanguages. En septembre 2003, iLoveLanguages offre 2.000 ressources linguistiques dans une centaine de langues. Quant à l'Internet Dictionary Project, Tyler Chambers y met fin en janvier 2007, faute de temps, tout en laissant les dictionnaires existants tels quels pour consultation ou téléchargement.

## 1995 > NetGlos, glossaire de l'internet

### **[Résumé]**

NetGlos - abrégé de «Multilingual Glossary of Internet Terminology» - est lancé en 1995 à l'initiative du WorldWide Language Institute (WWLI), un institut enseignant les langues via l'internet. NetGlos est un projet collaboratif en treize langues (allemand, anglais, chinois, croate, espagnol, français, grec, hébreu, hollandais/flamand, italien, maori, norvégien, portugais), avec la participation de nombreux traducteurs volontaires et autres professionnels des langues dans le monde entier. Comme expliqué en septembre 1998 par Brian King, directeur du WorldWide Language Institute, «avant qu'un nouveau terme ne soit accepté comme le terme correct, il y a une période d'instabilité avec plusieurs candidats en compétition. Souvent un terme emprunté à l'anglais est le point de départ et, dans de nombreux cas, il est aussi le point d'arrivée. Finalement émerge un vainqueur qui est ensuite utilisé aussi bien dans les dictionnaires techniques que dans le vocabulaire quotidien de l'utilisateur non spécialiste.»

\*\*\*

NetGlos - abrégé de «Multilingual Glossary of Internet Terminology» - est un glossaire coopératif lancé en 1995 à l'initiative du WorldWide Language Institute (WWLI), un institut enseignant les langues via l'internet.

Trois ans plus tard, NetGlos est disponible en treize langues (allemand, anglais, chinois, croate, espagnol, français, grec, hébreu, hollandais/flamand, italien, maori, norvégien, portugais), avec la participation de nombreux traducteurs et autres professionnels des langues dans le monde entier.

Brian King, directeur du WorldWide Language Institute, explique en septembre 1998: «Une grande partie de la terminologie technique disponible sur le web n'est pas encore traduite dans d'autres langues [que l'anglais]. Et, comme nous nous en sommes rendus compte dans NetGlos, la traduction de ces termes n'est pas toujours facile. Avant qu'un nouveau terme ne soit accepté comme le terme correct, il y a une période d'instabilité avec plusieurs candidats en compétition. Souvent un terme emprunté à l'anglais est le point de départ et, dans de nombreux cas, il est aussi le point d'arrivée. Finalement émerge un vainqueur qui est ensuite utilisé aussi bien dans les dictionnaires techniques que dans le vocabulaire quotidien de l'utilisateur non spécialiste. La dernière version de NetGlos est la version russe, et elle devrait être disponible dans deux semaines environ [à savoir fin septembre 1998]. Elle sera sans nul doute un excellent exemple du processus dynamique en cours pour la russification de la terminologie du web.»

Quelles sont les perspectives? «La technologie change à une allure frénétique. L'apprentissage durant toute la vie est une stratégie que nous devons tous adopter si nous voulons rester en tête et être compétitifs. C'est une tâche qui est déjà assez difficile dans un environnement anglophone. Si nous ajoutons à cela la complexité apportée par la communication dans un cyberspace multilingue et multiculturel, la tâche devient encore plus astreignante. Probablement davantage encore que par le passé, la coopération est aussi indispensable que la concurrence. Les germes

d'une coopération par le biais de l'internet existent déjà. Notre projet NetGlos dépend du bon vouloir de traducteurs volontaires dans de nombreux pays: Canada, États-Unis, Autriche, Norvège, Belgique, Israël, Portugal, Russie, Grèce, Brésil, Nouvelle-Zélande, etc. À mon avis, les centaines de visiteurs qui consultent quotidiennement les pages de NetGlos constituent un excellent témoignage du succès de ce type de relations de travail. Les relations de coopération s'accroîtront encore à l'avenir, mais pas nécessairement sur la base du volontariat.»

## **1995 > Plusieurs langues sur notre écran**

### **[Résumé]**

Comment se débrouiller pour afficher plusieurs langues à l'écran? Yoshi Mikami, informaticien à Fujisawa (Japon), lance d'abord un site web en décembre 1995 avant d'écrire le premier livre sur le sujet en août 1997. Son site bilingue anglais-japonais «The Languages of the World by Computers and the Internet» est plus connu sous le nom de Kotoba Home Page ou Logos Home Page (à ne pas confondre avec le site de la société de traduction Logos, lancé deux ans plus tard). La Kotoba Home Page donne un bref historique de chaque langue, ses caractéristiques, sa phonétique, son jeu de caractères et son encodage informatique. Fort de cette expérience, Yoshi Mikami co-écrit ensuite (avec Kenji Sekine et Nobutoshi Kohara) le premier livre sur le sujet, «Pour un web multilingue», publié en japonais par les éditions O'Reilly en août 1997 avant d'être traduit l'année suivante en anglais, en allemand et en français.

\*\*\*

Comment se débrouiller pour afficher plusieurs langues à l'écran? Yoshi Mikami lance d'abord un site web en décembre 1995 avant de publier le premier livre au monde sur le sujet en août 1997.

Yoshi Mikami, qui travaille à l'époque comme informaticien chez Asia Info Network à Fujisawa (Japon), lance le site bilingue anglais-japonais «The Languages of the World by Computers and the Internet», plus connu sous le nom de Logos Home Page ou Kotoba Home Page. Fort de cette expérience, il co-écrit ensuite (avec Kenji Sekine et Nobutoshi Kohara) le premier livre sur le sujet, «Pour un web multilingue», publié en japonais par les éditions O'Reilly Media en août 1997 avant d'être traduit l'année suivante en anglais, en allemand et en français.

Yoshi raconte en décembre 1998: «Ma langue maternelle est le japonais. Comme j'ai suivi mes études de troisième cycle aux États-Unis et que j'ai travaillé dans l'informatique, je suis devenu bilingue japonais/anglais américain. J'ai toujours été intéressé par différentes langues et cultures, aussi j'ai appris le russe, le français et le chinois dans la foulée. À la fin de 1995, j'ai créé sur le web "The Languages of the World by Computers and the Internet" et j'ai tenté de donner - en anglais et en japonais - un bref historique des six principales langues utilisées sur l'internet, ainsi que les caractéristiques propres à chaque langue, sa phonétique, son jeu de caractères et son encodage informatique. Suite à l'expérience acquise, j'ai invité mes deux associés à écrire un livre sur la conception, la création et la présentation de

pages web multilingues, livre qui fut publié en août 1997 dans son édition japonaise, le premier livre au monde sur un tel sujet.»

Plus généralement, «il y a des milliers d'années de cela, en Égypte, en Chine et ailleurs, les gens étaient plus sensibles au fait de communiquer leurs lois et leurs réflexions non seulement dans une langue mais dans plusieurs. Dans notre monde moderne, chaque État a le plus souvent adopté une seule langue de communication. À mon avis, l'internet verra l'utilisation plus grande de langues différentes et de pages multilingues - et pas seulement une gravitation autour de l'anglais américain - et un usage plus créatif de la traduction informatique multilingue. 99% des sites web créés au Japon sont en japonais!»

## **1995 > Global Reach, pour localiser les sites web**

### **[Résumé]**

Bill Dunlap est le fondateur de Euro-Marketing Associates, une société de conseil en marketing qu'il lance en 1985 à Paris et San Francisco. En 1995, il restructure cette société en service de conseil en ligne dénommé Global Reach, le but étant de promouvoir les sites web des entreprises américaines dans d'autres pays, afin d'attirer plus de visiteurs, et donc d'augmenter les ventes. Cette méthode comprend la traduction d'un site web dans plusieurs langues, la promotion active du site, et enfin l'accroissement de la fréquentation locale au moyen de bandeaux publicitaires ciblés. Bill Dunlap explique en décembre 1998: «Promouvoir un site est aussi important que le créer, sinon plus. On doit être préparé à utiliser au moins autant de temps et d'argent à promouvoir son site qu'on en a passé à l'origine à le créer. Le programme Global Reach permet de promouvoir un site dans des pays non anglophones, afin d'atteindre une clientèle plus large... et davantage de ventes.»

\*\*\*

La société Global Reach se spécialise d'emblée dans l'internationalisation des sites web, tout comme leur localisation dans plusieurs langues.

Bill Dunlap fonde d'abord Euro-Marketing Associates, une société de conseil en marketing qu'il lance en 1985 à Paris et San Francisco. En 1995, il restructure cette société en service de conseil en ligne dénommé Global Reach, le but étant de promouvoir les sites web des entreprises américaines dans d'autres pays, afin d'attirer plus de visiteurs, et donc d'augmenter les ventes. Cette méthode comprend la traduction d'un site web dans plusieurs langues, la promotion active du site, et enfin l'accroissement de la fréquentation locale au moyen de bandeaux publicitaires ciblés.

Bill Dunlap explique en décembre 1998: «Il y a très peu de gens aux États-Unis qui sont intéressés de communiquer dans plusieurs langues. Pour la plupart, ils pensent encore que le monde entier parle anglais. Par contre, en Europe, les pays sont petits, si bien que, depuis des siècles, une perspective internationale est nécessaire. (...)

Depuis 1981, début de mon activité professionnelle, j'ai été impliqué dans la venue de sociétés américaines en Europe. Ceci est pour beaucoup un problème de langue, puisque leurs informations commerciales doivent être disponibles dans les langues européennes pour être prises en compte en Europe. Comme le web est devenu populaire en 1995, j'ai donné à ces activités une dimension "en ligne", et j'en suis venu à promouvoir le cybercommerce européen auprès de mes compatriotes américains. (...)

Promouvoir un site est aussi important que le créer, sinon plus. On doit être préparé à utiliser au moins autant de temps et d'argent à promouvoir son site qu'on en a passé à l'origine à le créer. Le programme Global Reach permet de promouvoir un site dans des pays non anglophones, afin d'atteindre une clientèle plus large... avec davantage de ventes. Une société a de nombreuses bonnes raisons de considérer sérieusement le marché international. Global Reach est pour elle le moyen d'étendre son site web à de nombreux pays, de le présenter à des visiteurs en ligne dans leur propre langue, et de pénétrer le réseau de commerce en ligne présent dans ces pays.»

Bill Dunlap ajoute en juillet 1999: «Une fois que la page d'accueil d'un site est disponible en plusieurs langues, l'étape suivante est le développement du contenu dans chaque langue. Un webmestre notera quelles langues attirent plus de visiteurs (et donc plus de ventes) que d'autres. Ce seront dans ces langues que débutera une campagne de promotion multilingue sur le web. Parallèlement, il est toujours bon de continuer à augmenter le nombre de langues dans lesquelles un site web est disponible. Au début, seule la page d'accueil traduite en plusieurs langues suffit, mais ensuite il est souhaitable de développer un véritable secteur pour chaque langue.»

## **1996 > OneLook Dictionaries, point d'accès commun**

### **[Résumé]**

Robert Ware lance en avril 1996 le site OneLook Dictionaries en tant que point d'accès commun pour permettre une recherche rapide dans des centaines de dictionnaires généraux et spécialisés (affaires, informatique et internet, médecine, religion, sciences et techniques, sports, argot). Il explique en septembre 1998: «À titre personnel, je suis presque uniquement en contact avec des gens qui parlent une seule langue. (...) Être en contact avec le monde entier change cette approche des choses. Et la change en mieux ! (...) J'ai été long à inclure des dictionnaires non anglophones, en partie parce que je suis monolingue. Mais vous en trouverez maintenant quelques-uns.» OneLook Dictionaries compte 2 millions de termes provenant de 425 dictionnaires en 1998, 2,5 millions de termes provenant de 530 dictionnaires en 2000, 5 millions de termes provenant de 910 dictionnaires en 2003 et 19 millions de termes provenant de 1.060 dictionnaires en 2010.

\*\*\*

Le site OneLook Dictionaries est lancé en avril 1996 par Robert Ware pour offrir un point d'accès commun à des centaines de dictionnaires en ligne.

Outre les dictionnaires généraux, ces dictionnaires concernent l'économie et les affaires, l'informatique et l'internet, la médecine, la religion, les sciences et techniques, le sport et l'argot.

Robert Ware explique en septembre 1998: «À titre personnel, je suis presque uniquement en contact avec des gens qui parlent une seule langue et n'ont pas beaucoup de motivation pour développer leurs aptitudes linguistiques. Être en contact avec le monde entier change cette approche des choses. Et la change en mieux ! (...) J'ai été long à inclure des dictionnaires non anglophones, en partie parce que je suis monolingue. Mais vous en trouverez maintenant quelques-uns.»

Robert raconte aussi dans le même entretien: «Un fait intéressant s'est produit par le passé qui a été très instructif pour moi. En 1994, je travaillais pour un établissement scolaire et j'essayais d'installer un logiciel sur un modèle d'ordinateur particulier. J'ai trouvé une personne qui était en train de travailler sur le même problème, et nous avons commencé à échanger des courriels. Soudain, cela m'a frappé... Le logiciel avait été écrit à quarante kilomètres de là, mais c'était une personne située à l'autre bout du monde qui m'aidait. Les distances et l'éloignement géographique n'importaient plus! Et bien, ceci est formidable, mais à quoi cela nous mène-t-il? Je ne puis communiquer qu'en anglais mais, heureusement, mon correspondant pouvait utiliser aussi bien l'anglais que l'allemand qui était sa langue maternelle. L'internet a supprimé une barrière, celle de la distance, mais il subsiste la barrière de la langue, bien réelle.

Il semble que l'internet propulse simultanément les gens dans deux directions différentes. L'internet, anglophone à l'origine, relie les gens dans le monde entier. Par là-même, il favorise une langue commune pour communiquer. Mais il suscite aussi des contacts entre des personnes de langue différente et permet ainsi de développer un intérêt plus grand pour le multilinguisme. Si une langue commune est appréciable, elle ne remplace en aucun cas la nécessité de plusieurs langues. L'internet favorise ainsi à la fois une langue commune et le multilinguisme, et ceci est un facteur qui aide à trouver des solutions. L'intérêt croissant pour les langues et le besoin qu'on en a stimulent de par le monde la création de cours de langues et d'instruments d'aide linguistique, et l'internet fournit la possibilité de les rendre disponibles rapidement et à bon marché.»

OneLook Dictionaries compte 2 millions de termes provenant de 425 dictionnaires en 1998, 2,5 millions de termes provenant de 530 dictionnaires en 2000, 5 millions de termes provenant de 910 dictionnaires en 2003 et 19 millions de termes provenant de 1 060 dictionnaires en 2010.

## **1997 > Un web anglophone à 82,3%**

### **[Résumé]**

L'internet est d'abord anglophone à pratiquement 100%, puisqu'il débute en 1974 aux États-Unis en tant que réseau reliant les agences gouvernementales, les universités et les centres de recherche, grâce aux investissements considérables

du gouvernement, avant de s'étendre à la communauté anglophone puis au monde entier. Vingt ans plus tard, Babel, initiative conjointe de l'Internet Society et d'Alis Technologies, mène la première étude relative à la répartition des langues sur le web. Daté de juin 1997 et disponible en sept langues, le «Palmarès des langues de la toile» donne les pourcentages de 82,3% pour l'anglais, 4% pour l'allemand, 1,6% pour le japonais, 1,5% pour le français, 1,1% pour l'espagnol, 1,1% pour le suédois et 1% pour l'italien. Trois ans plus tard, en été 2000, 78% des pages web sont en anglais, et 50% des usagers sont non anglophones.

\*\*\*

En 1997, le web est anglophone à 82,3%, d'après Babel, un projet conjoint de l'Internet Society et d'Alis Technologies pour contribuer à l'internationalisation de l'internet.

À ses débuts, l'internet est anglophone à pratiquement 100%, puisqu'il débute en 1974 aux États-Unis en tant que réseau reliant les agences gouvernementales, les universités et les centres de recherche, grâce aux investissements considérables du gouvernement, avant de s'étendre à la communauté anglophone puis au monde entier. Son développement rapide est favorisé par l'invention du web en 1990 par Tim Berners-Lee puis le lancement en novembre 1993 du premier navigateur Mosaic, ancêtre de Netscape.

«Vers la communication sur internet dans toutes les langues...», tel est le sous-titre de la page d'accueil de Babel, une initiative de l'Internet Society et d'Alis Technologies. Pour mémoire, rappelons que l'Internet Society est fondée en 1992 par Vinton Cerf pour favoriser le développement de l'internet et qu'Alis Technologies est une société spécialisée dans le traitement automatique des langues.

En 1997, le site plurilingue (anglais, allemand, espagnol, français, italien, portugais, suédois) de Babel propose deux grands secteurs pour sa partie francophone: (a) un secteur «langues» en trois parties: langues du monde, glossaire typographique et linguistique, Francophonie; (b) un secteur «internet et multilinguisme» en deux parties: développer votre site web multilingue, codage des écritures du monde.

Babel mène la première étude sur la répartition des langues sur le web et publie cette étude dans les sept langues du site. Disponible en ligne en juin 1997, le « Palmarès des langues de la toile » donne les pourcentages de 82,3% pour l'anglais, 4% pour l'allemand, 1,6% pour le japonais, 1,5% pour le français, 1,1% pour l'espagnol, 1,1% pour le suédois et 1% pour l'italien.

Ce pourcentage de 82,3% pour l'anglais peut s'expliquer par plusieurs facteurs: (a) l'usage de l'anglais en tant que principale langue d'échange internationale; (b) la création d'un grand nombre de sites web aux États-Unis et au Canada depuis les débuts du web en 1990; (c) une proportion d'usagers particulièrement forte en Amérique du Nord par rapport au reste du monde, les ordinateurs étant bien meilleur marché qu'ailleurs, tout comme la connexion à l'internet sous forme de forfait mensuel à prix modique.

Selon Global Reach, société spécialisée dans l'internationalisation et la localisation des sites web, les usagers non anglophones sont au nombre de 56 millions en juillet 1998, avec 22,4% d'usagers hispanophones, 12,3% d'usagers nippons, 14% d'usagers germanophones et 10% d'usagers francophones. Sur les 500 millions d'habitants que compte l'Europe, 15% sont de langue maternelle anglaise, 28% ne parlent pas l'anglais et 32% consultent des pages web en anglais.

D'après Randy Hobler, un consultant en marketing internet de logiciels et services de traduction interviewé en septembre 1998, «l'augmentation de pages web dans d'autres langues que l'anglais n'est pas seulement due au fait qu'il y ait plus de sites et d'usagers dans des pays non anglophones, mais elle est également due au fait que les sociétés et les organisations localisent davantage leurs sites web et au fait qu'on utilise davantage la traduction automatisée pour proposer des sites web à partir ou vers d'autres langues.»

Randy explique aussi: «Comme l'internet n'a pas de frontières nationales, les internautes s'organisent selon d'autres critères propres au médium. En termes de multilinguisme, vous avez des communautés virtuelles, par exemple ce que j'appelle les "nations de langues", tous ces internautes qu'on peut regrouper selon leur langue maternelle quel que soit leur lieu géographique. Ainsi la nation de la langue espagnole inclut non seulement les internautes d'Espagne et d'Amérique latine, mais aussi tous les Hispanophones vivant aux États-Unis, ou encore ceux qui parlent espagnol au Maroc.»

Si Randy donne l'exemple d'une «nation de langue» hispanophone répartie sur trois continents, la même remarque vaut pour la Francophonie, une communauté de langue française présente sur cinq continents, ou encore la Créolophonie, une communauté de langue créole présente non seulement dans les Caraïbes mais aussi à Paris, Montréal et New York.

Le pourcentage d'internautes habitant hors des États-Unis atteint les 50% en juillet 1999. Le pourcentage d'usagers non anglophones atteint lui aussi les 50% en été 2000. Selon Global Reach, ce pourcentage est de 52,5% en été 2001, 57% en décembre 2001, 59,8% en avril 2002, 64,4% en septembre 2003 (dont 34,9% d'Européens non anglophones et 29,4% d'Asiatiques) et 64,2% en mars 2004 (dont 37,9% d'Européens non anglophones et 33% d'Asiatiques).

## **1997 > Une liste de langues européennes minoritaires**

### ***[Résumé]***

En 1997, Caoimhín Ó Donnáile est professeur d'informatique à l'Institut Sabhal Mòr Ostaig, situé sur l'île de Skye, en Écosse. Il dispense ses cours en gaélique écossais. Il est également le webmestre du site de l'Institut, un site trilingue (gaélique écossais, gaélique irlandais, anglais) qui se trouve être la principale source d'information mondiale sur le gaélique écossais. Sur ce site, il tient à jour la page «European Minority Languages», une liste de langues européennes minoritaires elle aussi trilingue, avec classement par ordre alphabétique de langues et par famille linguistique.

\*\*\*

Caoimhín Ó Donnáile est professeur d'informatique à l'Institut Sabhal Mór Ostaig, situé sur l'île de Skye, en Écosse. Il dispense ses cours en gaélique écossais. Il est également le webmestre du site de l'Institut, un site trilingue (gaélique écossais, gaélique irlandais, anglais) qui se trouve être la principale source d'information mondiale sur le gaélique écossais. Sur ce site, il propose la page «European Minority Languages», une liste de langues européennes minoritaires elle aussi trilingue, avec classement par ordre alphabétique de langues et par famille linguistique.

Interviewé en août 1998, Caoimhín détermine quatre points importants pour un internet plurilingue: «(a) L'internet a contribué et contribuera au développement fulgurant de l'anglais comme langue mondiale. (b) L'internet peut aussi grandement aider les langues minoritaires. Ceci ne se fera pas tout seul, mais seulement si les gens choisissent de défendre une langue. (c) Le web est très utile pour dispenser des cours de langues, et la demande est grande. (d) La norme Unicode (ISO 10646) pour les jeux de caractères est très importante et elle va grandement favoriser le multilinguisme sur le web.»

Qu'en est-il du gaélique? Caoimhín explique en mai 2001: «Nos étudiants utilisent un correcteur d'orthographe en gaélique et une base terminologique en ligne en gaélique. (...) Il est maintenant possible d'écouter la radio en gaélique (écossais et irlandais) en continu sur l'internet partout dans le monde. Une réalisation particulièrement importante a été la traduction en gaélique du navigateur Opera. C'est la première fois qu'un logiciel de cette taille est disponible en gaélique.»

Qu'en est-il des langues menacées? D'après Caoimhín, «l'internet accélère les choses dans les deux sens. Si les gens ne se soucient pas de préserver ces langues, l'internet et la mondialisation qui l'accompagne accéléreront considérablement leur disparition. Si les gens se soucient vraiment de les préserver, l'internet constituera une aide irremplaçable.»

## **1997 > Une base terminologique européenne**

### **[Résumé]**

Géré par le service de traduction de la Commission européenne et disponible sur le web depuis 1997, Eurodicautom est une base terminologique multilingue de termes économiques, scientifiques, techniques et juridiques permettant de combiner entre elles les onze langues officielles de l'Union européenne (allemand, anglais, danois, espagnol, finnois, français, grec, hollandais, italien, portugais, suédois), ainsi que le latin, avec une moyenne de 120.000 consultations par jour en 2003. Fin 2003, Eurodicautom annonce son intégration dans une base terminologique plus vaste regroupant le contenu de toutes les bases de l'Union européenne et qui serait disponible dans une vingtaine de langues du fait de l'élargissement de l'Union européenne prévu l'année suivante. La base IATE (InterActive Terminology for Europe) est lancée en interne au printemps 2004 puis en externe sur le web en juin 2007.

\*\*\*

Eurodicautom est une base terminologique multilingue de termes économiques, scientifiques, techniques et juridiques permettant de combiner entre elles les onze langues officielles de l'Union européenne et le latin.

Eurodicautom est l'oeuvre du Service de traduction de la Commission européenne. Initialement conçue pour aider les traducteurs en interne, la base terminologique est disponible sur le web en 1997 avec accès libre et gratuit pour pouvoir être utilisée par les fonctionnaires de l'Union européenne et les professionnels de langues de par le monde. Les onze langues officielles (allemand, anglais, danois, espagnol, finnois, français, grec, hollandais, italien, portugais, suédois) permettent de répondre aux besoins des 15 pays membres.

Un projet de base terminologique plus vaste est évoqué dès 1999 pour fusionner le contenu de toutes les bases terminologiques de l'Union européenne et renforcer ainsi la coopération inter-institutionnelle. Outre la Commission européenne, les partenaires de ce projet sont le Parlement européen, le Conseil de l'Union européenne, la Cour de justice, la Cour des comptes européenne, le Comité économique et social européen, le Comité des régions, la Banque européenne d'investissement, la Banque centrale européenne et le Centre de traduction des organes de l'Union européenne.

Fort de ses 120.000 consultations par jour, Eurodicautom annonce fin 2003 sa fermeture provisoire et son intégration dans une base plus vaste qui ne comporterait plus douze langues mais une vingtaine de langues, du fait de l'élargissement de l'Union européenne prévu l'année suivante (avec 25 pays membres en mai 2004 et 27 pays membres en janvier 2007).

La nouvelle base IATE (InterActive Terminology for Europe) voit le jour au printemps 2004, d'abord pour un usage interne au sein des institutions de l'Union européenne, puis pour un usage externe libre et gratuit sur le web en juin 2007, avec 1,4 million d'entrées dans les 23 langues officielles de l'Union européenne (allemand, anglais, bulgare, danois, espagnol, estonien, finnois, français, grec, hongrois, irlandais, italien, letton, lituanien, maltais, néerlandais, polonais, portugais, roumain, slovaque, slovène, suédois, tchèque), plus le latin.

Le site web de IATE est administré par le Centre de traduction des organes de l'Union européenne à Luxembourg (capitale du pays du même nom) pour le compte des partenaires du projet. Comme expliqué dans la brochure de IATE, elle-même disponible en 23 langues, «les termes sont introduits dans la base de données par les terminologues et les traducteurs de l'Union européenne sur la base des informations fournies par les traducteurs, les administrateurs, les juristes-linguistes, les experts et d'autres sources fiables.» Selon la même brochure, IATE comprend 8,4 millions de termes en 2010, dont 540.000 abréviations et 130.000 expressions.

## 1997 > Babel Fish, logiciel de traduction gratuit

### **[Résumé]**

En décembre 1997, AltaVista est le premier moteur de recherche à lancer un service gratuit de traduction automatisée de l'anglais vers cinq autres langues (allemand, espagnol, français, italien, portugais) et vice versa, la page web originale et la traduction apparaissant en vis-à-vis à l'écran. Appelé aussi AltaVista Translation, Babel Fish est l'oeuvre de Systran, société pionnière dans le traitement automatique des langues. Babel Fish est alimenté par des dictionnaires multilingues comprenant 2,5 millions de termes. Bien qu'ayant ses limites, avec un texte traduit très approximatif, ce service est immédiatement plébiscité par les 12 millions d'utilisateurs que compte la toile de l'époque, dont un nombre croissant d'utilisateurs non anglophones, et contribue grandement au plurilinguisme du web.

\*\*\*

En décembre 1997, le moteur de recherche AltaVista lance un logiciel de traduction gratuit de l'anglais vers cinq autres langues, dénommé Babel Fish ou AltaVista Translation, ce qui constitue une première sur le web.

À cette date, l'annuaire Yahoo! propose déjà une interface en sept langues (anglais, allemand, coréen, français, japonais, norvégien, suédois) pour prendre en compte un nombre croissant d'utilisateurs non anglophones. Le classement des sites en 63 sections est plus pointu que celui d'AltaVista, où ces tâches sont entièrement automatisées. Lorsqu'une recherche ne donne pas de résultat dans Yahoo!, elle est automatiquement aiguillée vers AltaVista, et réciproquement.

Babel Fish peut traduire une page web de l'anglais vers cinq autres langues (allemand, espagnol, français, italien, portugais) et vice versa, la page web originale et la traduction apparaissant en vis-à-vis à l'écran. On peut également traduire n'importe quel texte court en faisant un «copier-coller». Bien qu'ayant ses limites avec un texte traduit très approximatif, ce service est immédiatement plébiscité par les 12 millions d'utilisateurs que compte la toile de l'époque, dont un nombre croissant d'utilisateurs non anglophones, et contribue grandement au plurilinguisme de l'internet.

Alimenté par des dictionnaires multilingues comprenant 2,5 millions de termes, Babel Fish est l'oeuvre de Systran, société pionnière dans le traitement automatique des langues. Selon le site web de Systran, «un logiciel de traduction automatique traduit une langue naturelle dans une autre langue naturelle. La traduction automatique prend en compte la structure grammaticale de chaque langue et elle utilise des règles pour transférer la structure grammaticale de la langue source (texte à traduire) vers la langue cible (texte traduit). La traduction automatique ne remplace pas et n'est pas destinée à remplacer le traducteur humain.»

L'EAMT (European Association for Machine Translation) propose pour sa part la définition suivante sur son site: «La traduction automatique est l'utilisation de l'ordinateur pour la traduction de textes d'une langue naturelle à une autre. Elle fut un des premiers domaines de recherche en informatique. Il s'est avéré que cet objectif était difficile à atteindre. Cependant il existe aujourd'hui un certain

nombre de systèmes produisant un résultat qui, s'il n'est pas parfait, est de qualité suffisante pour être utile dans certaines applications spécifiques, en général dans le domaine de la documentation technique. De plus, les logiciels de traduction, qui sont essentiellement destinés à aider le traducteur humain à produire des traductions, jouissent d'une popularité croissante auprès d'organismes professionnels de traduction.»

D'autres logiciels de traduction automatique seront ensuite développés par Alis Technologies, Lernout & Hauspie, Globalink et Softissimo, avec des versions payantes et/ou gratuites disponibles sur le web. Quant à Babel Fish, il déménagera sur le site de Yahoo! en mai 2008.

## **1997 > Les outils de la société de traduction Logos**

### **[Résumé]**

En décembre 1997, la société de traduction Logos, basée à Modène, en Italie, avec des services en 35 langues, décide de mettre ses outils professionnels en accès libre sur le web pour en faire bénéficier la communauté internet. Le Logos Dictionary est un dictionnaire multilingue de 7,5 millions d'entrées (en septembre 1998). La Wordtheque est une base de données multilingue de 328 millions de termes, constituée à partir de milliers de traductions, notamment des romans et des documents techniques, avec une recherche possible par langue, par mot, par auteur ou par titre. Linguistic Resources offre un point d'accès unique à 553 glossaires. L'Universal Conjugator propose des tableaux de conjugaison dans 17 langues. Dix ans plus tard, en 2007, la Wordtheque (devenue la Logos Library) comprend 710 millions de termes, Linguistic Resources (qui n'a pas changé de nom) offre un point d'accès unique à 1.215 glossaires et Conjugation of Verbs (devenu l'Universal Conjugator) propose des tableaux de conjugaison dans 36 langues.

\*\*\*

Fin 1997, la société de traduction Logos décide de mettre ses outils professionnels en accès libre sur le web pour en faire bénéficier la communauté internet.

Fondé en 1979 par Rodrigo Vergara à Modène (Italie), Logos propose des services de traduction dans 35 langues en 1997, avec 300 traducteurs travaillant sur place et un réseau mondial de 2.500 traducteurs travaillant en free-lance. La moyenne de production est de 200 textes par jour.

Le principal outil professionnel en accès libre est le Logos Dictionary, un dictionnaire multilingue de 7,5 millions d'entrées (en septembre 1998). La Wordtheque est une base de données multilingue de 328 millions de termes, constituée à partir de milliers de traductions, notamment des traductions de romans et de documents techniques, avec une recherche possible par langue, mot, auteur ou titre. Linguistic Resources offre un point d'accès unique à 553 glossaires. L'Universal Conjugator propose des tableaux de conjugaison dans 17 langues.

Interviewé par Annie Kahn dans «Les mots pour le dire», un article du quotidien Le Monde daté du 7 décembre 1997, Rodrigo Vergara relate: «Nous voulions que nos traducteurs aient tous accès aux mêmes outils de traduction. Nous les avons donc mis à leur disposition sur internet, et tant qu'à faire nous avons ouvert le site au public. Cela nous a rendus très populaires, nous a fait beaucoup de publicité. L'opération a drainé vers nous de nombreux clients, mais aussi nous a permis d'étoffer notre réseau de traducteurs grâce aux contacts établis à la suite de cette initiative.»

Annie Kahn, auteure de l'article, explique: «Le site de Logos est beaucoup plus qu'un dictionnaire ou qu'un répertoire de liens vers d'autres dictionnaires en ligne. L'un des piliers du système est un logiciel de recherche documentaire fonctionnant sur un corpus de textes littéraires disponibles gratuitement sur internet. Lorsque l'on recherche la définition ou la traduction d'un mot, "didactique" par exemple, on trouve non seulement le résultat recherché, mais aussi une phrase d'une oeuvre littéraire utilisant ce mot (en l'occurrence, un essai de Voltaire). Un simple clic permet d'accéder au texte intégral de l'oeuvre ou de commander le livre grâce à un partenariat avec Amazon.com, le libraire en ligne bien connu. Il en est de même avec les traductions étrangères. Si aucun texte utilisant ce mot n'a été trouvé, le système fonctionne alors comme un moteur de recherche et renvoie aux sites web concernant ce mot. Pour certains termes, il est proposé d'en entendre la prononciation. Si une traduction manque, le système fait un appel au peuple. À chacun d'enrichir la base, les traducteurs de l'entreprise valident ensuite les traductions proposées.»

Dix ans plus tard, en 2007, la Wordtheque (devenue la Logos Library) comprend 710 millions de termes, Linguistic Resources (qui n'a pas changé de nom) offre un point d'accès unique à 1.215 glossaires et Conjugation of Verbs (devenu l'Universal Conjugator) propose des tableaux de conjugaison dans 36 langues.

## **1997 > Des bases terminologiques spécialisées**

### **[Résumé]**

En 1997 et 1998, des organisations internationales mettent leurs bases terminologiques spécialisées en accès libre sur le web, par exemple la base ILOTERM de l'Organisation internationale du Travail (OIT), la base TERMITE de l'Union internationale des télécommunications (UIT) et la base WHOTERM de l'Organisation mondiale de la santé (OMS), ce qui leur permet d'être à la disposition des traducteurs et linguistiques du monde entier. Prenons l'exemple de la base terminologique de l'OMS. WHOTERM, acronyme de «WHO Terminology Information System», est une base terminologique trilingue (anglais, espagnol, français) dont le but est d'«améliorer la rigueur et la cohérence des textes rédigés, préparés ou traduits. Elle permet également à tous ceux qui collaborent à des programmes techniques de l'OMS d'enrichir les terminologies nouvelles, de promouvoir leur normalisation et de garantir leur diffusion.»

\*\*\*

En 1997 et 1998, des organisations internationales mettent leurs bases terminologiques spécialisées en accès libre sur le web, ce qui leur permet d'être à la disposition des traducteurs et linguistiques du monde entier.

C'est le cas par exemple pour la base ILOTERM de l'Organisation internationale du Travail (OIT), la base TERMITE de l'Union internationale des télécommunications (UIT) et la base WHOTERM de l'Organisation mondiale de la santé (OMS).

ILOTERM est une base terminologique quadrilingue (allemand, anglais, espagnol, français) qui est l'oeuvre de l'Unité de terminologie et de références du Service des documents officiels (OFFDOC) de l'Organisation internationale du Travail (OIT). Comme indiqué sur le site web en 1998, «sa principale finalité est d'apporter des solutions, conformes à l'usage courant, à des problèmes terminologiques dans le domaine du travail et des questions sociales. Les termes figurent en anglais avec leurs équivalents en français, espagnol et/ou allemand. La base de données contient également (dans une à quatre langues) des articles concernant la structure et les programmes de l'OIT, les noms officiels d'institutions internationales, d'organismes nationaux et d'organisations nationales d'employeurs et de travailleurs, ainsi que les titres de réunions et d'instruments internationaux.»

TERMITE - acronyme de «Base de données terminologique des Télécommunications de l'UIT» - est également quadrilingue (anglais, espagnol, français, russe) et géré par la Section de traduction de l'Union internationale des télécommunications (UIT). Comme indiqué sur le site web, «TERMITE contient tous les termes qui apparaissent dans tous les glossaires de l'UIT imprimés depuis 1980, ainsi que des termes plus récents en rapport avec les différentes activités de l'Union (en tout quelque 59.000 entrées). Normalement les collaborateurs qui s'occupent de l'amélioration et de la mise à jour de cette base de données sont des traducteurs ou des réviseurs techniques. TERMITE est surtout visité par les traducteurs internes mais aussi par des utilisateurs externes, travaillant dans le domaine des télécommunications.»

WHOTERM - acronyme de «WHO Terminology Information System» - est la base terminologique trilingue (anglais, espagnol, français) de l'Organisation mondiale de la santé (OMS), dont le but est d'«améliorer la rigueur et la cohérence des textes rédigés, préparés ou traduits. Elle permet également à tous ceux qui collaborent à des programmes techniques de l'OMS d'enrichir les terminologies nouvelles, de promouvoir leur normalisation et de garantir leur diffusion.»

## **1998 > La nécessité d'une «démocratie linguistique»**

### **[Résumé]**

Brian King, directeur du WorldWide Language Institute (WWLI), développe le concept de «démocratie linguistique» en septembre 1998: «Dans un rapport de l'UNESCO du début des années 1950, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur l'internet dans sa langue maternelle pourrait bien être son équivalent à l'Âge de l'Information. Si l'internet doit vraiment devenir le réseau mondial qu'on nous promet,

tous les usagers devraient y avoir accès sans problème de langue. Considérer l'internet 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.»

\*\*\*

Brian King, directeur du WorldWide Language Institute (WWLI), développe le concept de «démocratie linguistique» en septembre 1998.

Il explique lors d'un entretien par courriel: «Dans un rapport de l'UNESCO du début des années 1950, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur l'internet dans sa langue maternelle pourrait bien être son équivalent à l'Âge de l'Information. Si l'internet doit vraiment devenir le réseau mondial qu'on nous promet, tous les usagers devraient y avoir accès sans problème de langue. Considérer l'internet 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.»

Un facteur de développement d'un internet multilingue est «la compétition entre les grandes sociétés pour une part de marché global», avec «l'exportation des technologies de l'information dans le monde entier. La popularisation est maintenant effective à l'échelon mondial, et l'anglais n'est plus nécessairement la langue obligée de l'utilisateur. Il n'y a plus vraiment de langue indispensable, il y a les langues propres aux utilisateurs. 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 d'informatique. La demande des utilisateurs non anglophones - et l'effort entrepris par les sociétés de haute technologie se faisant concurrence pour obtenir les marchés mondiaux - ont fait de la localisation un secteur en expansion rapide dans le développement des logiciels et du matériel informatique.»

Un autre facteur est le développement du commerce électronique: «De même que l'utilisateur non anglophone peut maintenant avoir accès aux nouvelles technologies dans sa propre langue, l'impact du commerce électronique peut constituer une force majeure qui fasse du multilinguisme la voie la plus naturelle vers le cyberspace. Les vendeurs de produits et services dans le marché virtuel mondial que devient l'internet doivent être préparés à desservir un monde virtuel qui soit aussi multilingue que le monde physique. S'ils veulent réussir, ils doivent s'assurer qu'ils parlent bien la langue de leurs clients!»

C'est ce que fait Bill Dunlap, fondateur de Euro-Marketing Associates, une société de conseil en marketing qu'il lance en 1985 à Paris et San Francisco. En 1995, il restructure cette société en service de conseil en ligne dénommé Global Reach, le but étant de promouvoir en Europe les sites web des entreprises américaines, afin d'attirer plus de visiteurs, et donc d'augmenter les ventes. Cette méthode comprend la traduction d'un site web dans plusieurs langues (ce qu'on appelle la localisation d'un site), la promotion active des sites traduits et enfin l'accroissement de la fréquentation locale au moyen de bandeaux publicitaires ciblés.

Bill Dunlap écrit en décembre 1998: «Il y a très peu de gens aux États-Unis qui sont intéressés de communiquer dans plusieurs langues. Pour la plupart, ils pensent encore que le monde entier parle anglais. Par contre, en Europe, les pays sont petits, si bien que, depuis des siècles, une perspective internationale est nécessaire», d'où l'intérêt de son travail sur les deux continents.

Coordinateur d'ELSNET (European Network of Excellence in Human Language Technologies), Steven Krauwer explique en septembre 1998: «En tant que citoyen européen, je pense que le multilinguisme sur le web est absolument essentiel. À mon avis, ce n'est pas une situation saine à long terme que seuls ceux qui ont une bonne maîtrise de l'anglais puissent pleinement exploiter les bénéfices du web. En tant que chercheur (spécialisé dans la traduction automatique), je vois le multilinguisme comme un défi majeur: pouvoir garantir que l'information sur le web soit accessible à tous, indépendamment des différences de langue.»

Pour ce faire, il suggère plusieurs solutions pratiques: «(a) en ce qui concerne les auteurs: une meilleure formation des auteurs de sites web pour exploiter les combinaisons possibles permettant d'améliorer la communication en surmontant la barrière de la langue (et pas seulement par un vernis superficiel); (b) en ce qui concerne les usagers: des logiciels de traduction de type AltaVista Translation [Babel Fish], dont la qualité n'est pas frappante, mais qui a le mérite d'exister; (c) en ce qui concerne les logiciels de navigation: des logiciels de traduction intégrés, particulièrement pour les langues non dominantes, et des dictionnaires intégrés plus rapides à consulter.»

Si l'internet gagne le monde entier avec l'anglais comme principale langue d'échange, tout le monde ne parle pas l'anglais, loin de là, et même ceux qui le lisent préfèrent consulter le web dans leur propre langue. Pour élargir leur champ d'action, les sociétés doivent donc proposer des sites bilingues, trilingues sinon plurilingues, en adaptant leur contenu à un public spécifique, que ce soit un pays ou une communauté linguistique. D'où la nécessité de l'internationalisation et de la localisation des sites, qui devient indispensable dans les années qui suivent, avec les sociétés et organismes anglophones proposant leurs sites à la fois en anglais et dans d'autres langues, et les sociétés et organismes non anglophones proposant leurs sites dans leur(s) propre(s) langue(s) et en anglais.

## **1999 > Les dictionnaires bilingues de WordReference.com**

### **[Résumé]**

Michael Kellogg crée le site WordReference.com en 1999. Il raconte beaucoup plus tard sur son site: «J'ai débuté ce site en 1999 pour procurer des dictionnaires bilingues gratuits en ligne et d'autres outils pour tous sur l'internet. Depuis, le site s'est progressivement développé pour devenir l'un des sites de dictionnaires en ligne les plus utilisés, et le principal dictionnaire en ligne pour les paires de langues anglais-espagnol, anglais-français, anglais-italien, espagnol-français et espagnol-portugais. Ce site est toujours classé sans interruption parmi les 500 sites les plus visités du web.» WordReference.com offre également des forums linguistiques très actifs, tout comme des versions allégées de ses dictionnaires pour appareil mobile.

\*\*\*

Michael Kellogg crée le site WordReference.com en 1999 pour procurer des dictionnaires bilingues gratuits en ligne.

Il raconte beaucoup plus tard sur son site: «L'internet a été un incroyable outil ces dernières années pour rassembler des gens du monde entier. L'un des principaux obstacles à cela reste bien entendu la langue. Le contenu de l'internet est pour une grande part en anglais et de très nombreux usagers lisent ces pages alors que l'anglais est leur deuxième langue et non leur langue maternelle. De par mes propres expériences avec la langue espagnole, je sais que de nombreux lecteurs comprennent une grande partie de ce qu'ils lisent, mais pas la totalité.

J'ai débuté ce site en 1999 pour procurer des dictionnaires bilingues gratuits en ligne et d'autres outils pour tous sur l'internet. Depuis, le site s'est progressivement développé pour devenir l'un des sites de dictionnaires en ligne les plus utilisés, et le principal dictionnaire en ligne pour les paires de langues anglais-espagnol, anglais-français, anglais-italien, espagnol-français et espagnol-portugais. Ce site est toujours classé sans interruption parmi les 500 sites les plus visités du web. Aujourd'hui, je suis heureux de continuer à améliorer ces dictionnaires, les autres outils linguistiques du site et les forums de langues.»

En 2010, outre ces dictionnaires, WordReference.com propose un dictionnaire monolingue anglais ainsi que des dictionnaires de l'anglais vers d'autres langues (arabe, chinois, coréen, grec, japonais, polonais, portugais, roumain, tchèque, turc) et vice versa. Pour la langue espagnole, un dictionnaire monolingue voisine avec un dictionnaire de synonymes, un dictionnaire espagnol-français et un dictionnaire espagnol-portugais. Des tableaux de conjugaison sont disponibles pour l'espagnol, le français et l'italien. L'allemand et le russe disposent d'un dictionnaire monolingue. WordReference Mini est une version miniature du site pour intégration dans d'autres sites, par exemple des sites d'apprentissage de langues. Une version pour appareil mobile est disponible pour plusieurs dictionnaires, de l'anglais vers l'espagnol, le français et l'italien, et vice versa, avec d'autres paires de langues à venir.

## **1999 > L'internet, outil indispensable pour les traducteurs**

### **[Résumé]**

L'internet devient un outil important pour les traducteurs, tout comme «une source indispensable et inépuisable d'informations», comme expliqué par Marcel Grangier, responsable de la section française des Services linguistiques centraux de l'Administration fédérale suisse. Il écrit en janvier 1999: «Travailler sans internet est devenu tout simplement impossible. Au-delà de tous les outils et commodités utilisés (messagerie électronique, consultation de la presse électronique, activités de services au profit de la profession des traducteurs), internet reste pour nous une source indispensable et inépuisable d'informations dans ce que j'appellerais le "secteur non structuré" de la toile. Pour illustrer le propos, lorsqu'aucun site comportant de l'information organisée ne fournit de réponse à un problème de

traduction, les moteurs de recherche permettent dans la plus grande partie des cas de retrouver le chaînon manquant quelque part sur le réseau.» Le service de Marcel Grangier gère notamment le répertoire Dictionnaires électroniques, une liste exhaustive de dictionnaires disponibles en ligne.

\*\*\*

L'internet devient un outil important pour les traducteurs, tout comme «une source indispensable et inépuisable d'informations».

Marcel Grangier est responsable de la section française des Services linguistiques centraux de l'Administration fédérale suisse. Il explique en janvier 1999: «Travailler sans internet est devenu tout simplement impossible. Au-delà de tous les outils et commodités utilisés (messagerie électronique, consultation de la presse électronique, activités de services au profit de la profession des traducteurs), internet reste pour nous une source indispensable et inépuisable d'informations dans ce que j'appellerais le "secteur non structuré" de la toile. Pour illustrer le propos, lorsqu'aucun site comportant de l'information organisée ne fournit de réponse à un problème de traduction, les moteurs de recherche permettent dans la plus grande partie des cas de retrouver le chaînon manquant quelque part sur le réseau.»

D'après lui, «le multilinguisme sur internet peut être considéré comme une fatalité heureuse et surtout irréversible. C'est dans cette optique qu'il convient de creuser la tombe des rabat-joie dont le seul discours est de se plaindre d'une suprématie de l'anglais. La suprématie de l'anglais n'est pas un mal en soi, dans la mesure où elle résulte de réalités essentiellement statistiques (plus de PC par habitant, plus de locuteurs de cette langue, etc.). La riposte n'est pas de "lutter contre l'anglais" et encore moins de s'en tenir à des jérémiades, mais de multiplier les sites en d'autres langues. Notons qu'en qualité de service de traduction, nous préconisons également le multilinguisme des sites eux-mêmes. La multiplication des langues présentes sur l'internet est inévitable, et ne peut que bénéficier aux échanges multiculturels. Pour que ces échanges prennent place dans un environnement optimal, il convient encore de développer les outils qui amélioreront la compatibilité. La gestion complète des diacritiques ne constitue qu'un exemple de ce qui peut encore être entrepris.»

Géré par le service de Marcel Grangier, le répertoire Dictionnaires électroniques est liste très complète de dictionnaires monolingues (allemand, anglais, espagnol, français, italien), bilingues et multilingues disponibles en ligne, complétée par des répertoires d'abréviations et acronymes et des répertoires géographiques, essentiellement des atlas. Ce répertoire déménagera plus tard sur le nouveau site de la Conférence des Services de traduction des États européens (CST).

Marcel Grangier précise en janvier 2000: "Les Dictionnaires électroniques ne sont qu'une partie de l'ensemble [du site web], et d'autres secteurs documentaires ont trait à l'administration, au droit, à la langue française, etc., sans parler des informations générales. (...) Conçu d'abord comme un service intranet, notre site web se veut en premier lieu au service des traducteurs opérant en Suisse, qui souvent travaillent sur la même matière que les traducteurs de l'Administration fédérale, mais également, par certaines rubriques, au service de n'importe quel autre traducteur où qu'il se trouve.»

Maria Victoria Marinetti, traductrice de nationalité mexicaine, est titulaire d'un doctorat en ingénierie. Elle raconte en août 1999: «J'ai accès à un nombre important d'informations au niveau mondial, ce qui est très intéressant pour moi. J'ai également la possibilité de transmettre ou de recevoir des fichiers, dans un va-et-vient d'information constant. L'internet me permet de recevoir ou d'envoyer des traductions générales ou techniques du français vers l'espagnol et vice versa, ainsi que des textes espagnols corrigés. Dans le domaine technique ou chimique, je propose une aide technique, ainsi que des informations sur l'exportation d'équipes de haute technologie vers le Mexique ou d'autres pays d'Amérique latine.»

En ce qui concerne le multilinguisme, «il est très important de pouvoir communiquer en différentes langues. Je dirais même que c'est obligatoire, car l'information donnée sur l'internet est à destination du monde entier, alors pourquoi ne l'aurions-nous pas dans notre propre langue ou dans la langue que nous souhaitons utiliser? Information mondiale, mais pas de vaste choix dans les langues, ce serait contradictoire, pas vrai?»

En 2000, l'internet est multilingue, et la moitié des usagers n'est pas de langue maternelle anglaise, mais la barrière de la langue est loin d'avoir disparu. Si toutes les langues sont désormais représentées sur le web, on oublie trop souvent que de nombreux usagers sont unilingues, et que même les polyglottes ne peuvent connaître toutes les langues. Il importe aussi d'avoir à l'esprit l'ensemble des langues, et pas seulement les langues dominantes. Il reste à créer des passerelles entre les communautés linguistiques pour favoriser la circulation des informations dans d'autres langues, notamment en améliorant la qualité des logiciels de traduction.

## **1999 > La nécessité d'une information bilingue**

### **[Résumé]**

Henk Slettenhaar, professeur en technologies des communications à la Webster University de Genève (Suisse), insiste régulièrement sur la nécessité de sites d'information bilingues, dans la langue originale et en anglais. Il écrit en 1999 : «Les communautés locales présentes sur le web devraient en tout premier lieu utiliser leur langue pour diffuser des informations. Si elles veulent présenter ces informations à la communauté mondiale, celles-ci doivent être disponibles aussi en anglais. Je pense qu'il existe un réel besoin de sites bilingues. (...) À mon avis, il existe deux types de recherches sur le web. La première est la recherche globale dans le domaine des affaires et de l'information. Pour cela, la langue est d'abord l'anglais, avec des versions locales si nécessaire. La seconde, ce sont les informations locales de tous ordres dans les endroits les plus reculés. Si l'information est à destination d'une ethnie ou d'un groupe linguistique, elle doit d'abord être disponible dans la langue de l'ethnie ou du groupe, avec peut-être un résumé en anglais.»

\*\*\*

Le web étant un médium à destination du monde, il paraît indispensable d'offrir des informations bilingues sinon plurilingues.

Henk Slettenhaar est professeur à la Webster University de Genève (Suisse), où il enseigne les technologies des communications. Il insiste régulièrement sur la nécessité de sites d'information bilingues, dans la langue originale et en anglais. «Les communautés locales présentes sur le web devraient en tout premier lieu utiliser leur langue pour diffuser des informations, écrit-il en décembre 1998. Si elles veulent présenter ces informations à la communauté mondiale, celles-ci doivent être disponibles aussi en anglais. Je pense qu'il existe un réel besoin de sites bilingues. (...) Mais je suis enchanté qu'il existe maintenant tant de documents disponibles dans leur langue originale. Je préfère de beaucoup lire l'original avec difficulté plutôt qu'une traduction médiocre.»

Il ajoute en août 1999: «À mon avis, il existe deux types de recherches sur le web. La première est la recherche globale dans le domaine des affaires et de l'information. Pour cela, la langue est d'abord l'anglais, avec des versions locales si nécessaire. La seconde, ce sont les informations locales de tous ordres dans les endroits les plus reculés. Si l'information est à destination d'une ethnie ou d'un groupe linguistique, elle doit d'abord être disponible dans la langue de l'ethnie ou du groupe, avec peut-être un résumé en anglais.»

Guy Antoine crée en avril 1998 le site Windows on Haiti pour promouvoir la culture haïtienne et sa langue. Il croit toutefois en la nécessité de l'anglais en tant que langue commune et relate en novembre 1999: «Pour des raisons pratiques, l'anglais continuera de dominer le web. Je ne pense pas que ce soit une mauvaise chose, en dépit des sentiments régionalistes qui s'y opposent, parce que nous avons besoin d'une langue commune permettant de favoriser les communications à l'échelon international. L'internet peut héberger des informations utiles sur les langues minoritaires, qui seraient autrement amenées à disparaître sans laisser de traces. De plus, à mon avis, l'internet incite les gens à apprendre les langues associées aux cultures qui les intéressent. Ces personnes réalisent rapidement que la langue d'un peuple est un élément fondamental de sa culture. (...) Dans Windows on Haiti, la langue principale est l'anglais, mais on y trouve tout aussi bien un forum de discussion animé conduit en kreyòl [créole haïtien]. On y trouve également des documents sur Haïti en français et dans l'ancien créole colonial, et je suis prêt à publier d'autres documents en espagnol et dans diverses langues. Je ne propose pas de traductions, mais le multilinguisme est effectif sur ce site, et je pense qu'il deviendra de plus en plus la norme sur le web.»

Bakayoko Bourahima est responsable de la bibliothèque de l'ENSEA (École nationale supérieure de statistique et d'économie appliquée) d'Abidjan, en Côte d'Ivoire. Il écrit en juillet 2000: «Pour nous les Africains francophones, le diktat de l'anglais sur la toile représente pour la masse un double handicap d'accès aux ressources du réseau. Il y a d'abord le problème de l'alphabétisation qui est loin d'être résolu et que l'internet va poser avec beaucoup plus d'acuité, ensuite se pose le problème de la maîtrise d'une seconde langue étrangère et son adéquation à l'environnement culturel. (...) Nos systèmes éducatifs ont déjà beaucoup de mal à optimiser leurs performances, en raison, selon certains spécialistes, des contraintes de l'utilisation du français comme langue de formation de base. Il est donc de plus

en plus question de recourir aux langues vernaculaires pour les formations de base, pour "désenclaver" l'école en Afrique et l'impliquer au mieux dans la valorisation des ressources humaines. Comment faire? Je pense qu'il n'y a pas de chance pour nous de faire prévaloir une quelconque exception culturelle sur la toile, ce qui serait de nature tout à fait grégaire. Il faut donc que les différents blocs linguistiques s'investissent beaucoup plus dans la promotion de leur accès à la toile, sans oublier leurs différentes spécificités internes.»

Bruno Didier, webmestre de la bibliothèque de l'Institut Pasteur, écrit en août 1999: «Internet n'est une propriété ni nationale, ni linguistique. C'est un vecteur de culture, et le premier support de la culture, c'est la langue. Plus il y a de langues représentées dans leur diversité, plus il y aura de cultures sur internet. Je ne pense pas qu'il faille justement céder à la tentation systématique de traduire ses pages dans une langue plus ou moins universelle. Les échanges culturels passent par la volonté de se mettre à la portée de celui vers qui on souhaite aller. Et cet effort passe par l'appréhension de sa langue. Bien entendu c'est très utopique comme propos. Concrètement, lorsque je fais de la veille, je peste dès que je rencontre des sites norvégiens ou brésiliens sans un minimum d'anglais.»

Alain Bron, consultant en systèmes d'information et écrivain, explique en novembre 1999: «Il y aura encore pendant longtemps l'usage de langues différentes et tant mieux pour le droit à la différence. Le risque est bien entendu l'envahissement d'une langue au détriment des autres, donc l'aplanissement culturel. Je pense que des services en ligne vont petit à petit se créer pour pallier cette difficulté. Tout d'abord, des traducteurs pourront traduire et commenter des textes à la demande, et surtout les sites de grande fréquentation vont investir dans des versions en langues différentes, comme le fait l'industrie audiovisuelle.»

Luc dall'Armellina, co-auteur et webmestre d'oVosite, un espace d'écriture hypermédia, écrit en juin 2000: «L'anglais s'impose sans doute parce qu'il est devenu la langue commerciale d'échange généralisée; il semble important que toutes les langues puissent continuer à être représentées parce que chacune d'elle est porteuse d'une vision "singulière" du monde. La traduction simultanée (proposée par AltaVista par exemple) ou les versions multilingues d'un même contenu me semblent aujourd'hui les meilleures réponses au danger de pensée unique que représenterait une seule langue d'échange. Peut-être appartient-il aux éditeurs des systèmes d'exploitation (ou de navigateurs?) de proposer des solutions de traduction partielle, avec toutes les limites connues des systèmes automatiques de traduction...»

En été 2000, 50% des usagers de l'internet sont non anglophones. Le cap des 50% étant désormais franchi, il reste à diversifier les langues sur une toile dont 78% des pages sont encore en anglais.

## **2000 > Encyclopédies et dictionnaires en ligne**

### **[Résumé]**

Les premières encyclopédies de référence disponibles sur le web émanent de versions imprimées. C'est aussi le cas des dictionnaires en ligne. WebEncyclo est

mis en ligne par les éditions Atlas en décembre 1999 avec accès libre et gratuit, tout comme l'Encyclopaedia Universalis, mais avec accès payant. Le site Britannica.com est lui aussi mis en ligne à la même date pour proposer le contenu des 32 volumes de l'Encyclopaedia Britannica, d'abord en accès libre puis en accès payant. Les premiers dictionnaires de référence en ligne émanent eux aussi de versions imprimées. Le Dictionnaire universel francophone en ligne des éditions Hachette est disponible en accès libre dès 1997. Les 20 volumes de l'Oxford English Dictionary (OED) sont mis en ligne en mars 2000 avec accès payant. Conçu directement pour le web, le Grand dictionnaire terminologique (GDT) est mis en ligne en septembre 2000 avec accès libre et gratuit.

\*\*\*

Les premières encyclopédies de référence disponibles sur le web émanent de versions imprimées. C'est aussi le cas des dictionnaires en ligne.

### ***Les encyclopédies***

WebEncyclo (aujourd'hui disparu), publié par les éditions Atlas, est la première grande encyclopédie francophone en accès libre, avec mise en ligne en décembre 1999. La recherche est possible par mots-clés, thèmes, médias (à savoir les cartes, liens internet, photos et illustrations) et idées. Un appel à contribution incite les spécialistes d'un sujet donné à envoyer des articles, qui sont regroupés dans la section «WebEncyclo contributif». Après avoir été libre, l'accès est ensuite soumis à une inscription préalable gratuite.

La version web de l'Encyclopaedia Universalis est elle aussi mise en ligne en décembre 1999, ce qui représente un ensemble de 28.000 articles signés de 4.000 auteurs. Si la consultation est payante sur la base d'un abonnement annuel, de nombreux articles sont également en accès libre.

Le site Britannica.com est mis en ligne à la même date, en tant que première grande encyclopédie anglophone en accès libre. Le site propose l'équivalent numérique des 32 volumes de l'Encyclopaedia Britannica (15e édition), en complément de la version imprimée et de la version CD-Rom, toutes deux payantes. Le site web offre aussi une sélection d'articles issus de 70 magazines, un guide des meilleurs sites, un choix de livres, etc., le tout étant accessible à partir d'un moteur de recherche unique. En septembre 2000, le site fait partie des cent sites les plus visités du web. En juillet 2001, la consultation devient payante sur la base d'un abonnement annuel ou mensuel. Beaucoup plus tard, en 2009, Britannica.com ouvre son site à des contributeurs externes, avec inscription obligatoire pour écrire et modifier des articles.

### ***Les dictionnaires***

Le premier grand dictionnaire de langue française en accès libre est le Dictionnaire universel francophone en ligne, qui répertorie 45.000 mots et 116.000 définitions tout en présentant «sur un pied d'égalité, le français dit "standard" et les mots et expressions en français tel qu'on le parle sur les cinq continents». Issu de la collaboration entre Hachette et l'AUPELF-UREF (devenu depuis l'AUF - Agence

universitaire de la Francophonie), il est mis en ligne dès 1997 et correspond à la partie «noms communs» du dictionnaire imprimé disponible chez Hachette.

L'équivalent pour la langue anglaise est le site Merriam-Webster OnLine, qui donne librement accès au Collegiate Dictionary, au Collegiate Thesaurus et à d'autres outils linguistiques.

En mars 2000, les 20 volumes de l'Oxford English Dictionary (OED) sont mis en ligne par l'Oxford University Press (OUP). La consultation du site est payante. Le dictionnaire bénéficie d'une mise à jour trimestrielle d'environ 1.000 entrées nouvelles ou révisées.

En mars 2002, deux ans après cette première expérience, l'Oxford University Press lance l'Oxford Reference Online (ORO), une vaste encyclopédie conçue cette fois directement pour le web et consultable elle aussi sur abonnement payant. Avec 60.000 pages et un million d'entrées, elle représente l'équivalent d'une centaine d'ouvrages de référence.

### ***Un dictionnaire bilingue***

Conçu lui aussi directement pour le web, avec accès libre et gratuit, le Grand dictionnaire terminologique (GDT) est un dictionnaire bilingue français-anglais de trois millions de termes appartenant au vocabulaire industriel, scientifique et commercial. Sa mise en ligne en septembre 2000 est le résultat d'un partenariat entre l'Office québécois de la langue française (OQLF), auteur du dictionnaire, et Semantix, société spécialisée dans les solutions logicielles linguistiques.

Dès le premier mois, le GDT est consulté par 1,3 million de personnes, avec 60.000 requêtes par jour. La gestion du GDT est ensuite assurée par Convera Canada, avec 3,5 millions de requêtes mensuelles en février 2003. Une nouvelle version du GDT est mise en ligne en mars 2003. La gestion du dictionnaire est désormais assurée par l'OQLF lui-même, et non plus par une société prestataire, avec l'ajout du latin comme troisième langue.

## **2000 > Le portail yourDictionary.com**

### ***[Résumé]***

Robert Beard, professeur de langues à l'Université Bucknell (États-Unis), crée d'abord en 1995 le site «A Web of Online Dictionaries» (800 liens en septembre 1998), qui est un répertoire de dictionnaires en ligne dans diverses langues, auquel s'ajoutent d'autres sections: dictionnaires multilingues, dictionnaires anglophones spécialisés, thesaurus, vocabulaires, grammaires, glossaires et méthodes de langues. Une section appelée Linguistic Fun propose des outils linguistiques pour non spécialistes. Robert Beard cofonde ensuite le portail yourDictionary.com en y intégrant son site précédent, avec mise en ligne en février 2000. Le portail répertorie 1.800 dictionnaires dans 250 langues en septembre 2003, et 2.500 dictionnaires dans 300 langues en avril 2007. Soucieux de servir toutes les langues sans exception, le portail propose aussi l'Endangered Language Repository, une section

spécifique consacrée aux langues menacées.

\*\*\*

Professeur de langues à la Bucknell University (États-Unis), Robert Beard cofonde en février 2000 le portail yourDictionary.com, un portail pour toutes les langues sans exception qui intègre son premier site, «A Web of Online Dictionaries», créé en 1995.

«A Web of Online Dictionaries» est un répertoire de dictionnaires en ligne dans diverses langues, auquel s'ajoutent d'autres sections: dictionnaires multilingues, dictionnaires anglophones spécialisés, thesaurus, vocabulaires, grammaires, glossaires et méthodes de langues, sans oublier une section proposant des outils linguistiques pour non spécialistes, sous le nom de Linguistic Fun.

Robert Beard écrit en septembre 1998: «On a d'abord craint que le web représente un danger pour le multilinguisme, étant donné que le langage HTML et d'autres langages de programmation sont basés sur l'anglais et qu'on trouve tout simplement plus de sites web en anglais que dans toute autre langue. Cependant, le site web que je gère montre que le multilinguisme est très présent et que le web peut en fait permettre de préserver des langues menacées de disparition. Je propose maintenant des liens vers des dictionnaires dans 150 langues différentes et des grammaires dans 65 langues différentes. De plus, comme les concepteurs de logiciels de navigation manifestent une attention nouvelle pour la diversité des langues dans le monde, ceci va encourager la présence de davantage encore de sites web dans différentes langues.»

Cinq ans après le lancement de son premier site, Robert Beard cofonde le portail yourDictionary.com, qui intègre son site précédent, avec mise en ligne du portail en février 2000. Il explique en janvier 2000 : «Nos nouvelles idées sont nombreuses. Nous projetons de travailler avec l'Endangered Language Fund aux États-Unis et en Grande-Bretagne pour rassembler des fonds pour cette fondation et nous publierons les résultats sur notre site. Nous aurons des groupes de discussion et des bulletins d'information sur les langues. Il y aura des jeux de langue destinés à se distraire et à apprendre les bases de la linguistique. La page Linguistic Fun deviendra un journal en ligne avec des extraits courts, intéressants et même amusants dans différentes langues, choisis par des experts du monde entier.»

Soucieux de servir toutes les langues sans exception, le portail propose l'Endangered Language Repository, une section spécifique consacrée aux langues menacées. «Les langues menacées sont essentiellement des langues non écrites, explique Robert Beard à la même date. Un tiers seulement des quelque 6.000 langues existant dans le monde sont à la fois écrites et parlées. Je ne pense pourtant pas que le web va contribuer à la perte de l'identité des langues et j'ai même le sentiment que, à long terme, il va renforcer cette identité. Par exemple, de plus en plus d'Indiens d'Amérique contactent des linguistes pour leur demander d'écrire la grammaire de leur langue et de les aider à élaborer des dictionnaires. Pour eux, le web est un instrument à la fois accessible et très précieux d'expression culturelle.»

En septembre 2003, yourDictionary.com répertorie 1.800 dictionnaires dans 250 langues, ainsi que d'autres outils linguistiques: vocabulaires, grammaires, glossaires, méthodes de langues, etc. En avril 2007, 2.500 dictionnaires et grammaires sont disponibles dans 300 langues.

Robert Beard écrivait de manière prémonitoire dès septembre 1998: «Le web sera une encyclopédie du monde faite par le monde pour le monde. Il n'y aura plus d'informations ni de connaissances utiles qui ne soient pas disponibles, si bien que l'obstacle principal à la compréhension internationale et interpersonnelle et au développement personnel et institutionnel sera levé. Il faudrait une imagination plus débordante que la mienne pour prédire l'effet de ce développement sur l'humanité.»

## **2000 > Le Projet Gutenberg et les langues**

### **[Résumé]**

Le Projet Gutenberg est un projet visionnaire lancé en juillet 1971 par Michael Hart pour créer des versions électroniques gratuites d'oeuvres littéraires et les diffuser dans le monde entier. En 2010, le Projet Gutenberg compte plus de 33.000 ebooks de grande qualité ainsi que des dizaines de milliers de téléchargements par jour. Il dispose de sites web aux États-Unis, en Australie, en Europe et au Canada, avec 40 sites miroirs répartis sur toute la planète. Les livres numériques sont surtout en anglais, mais le multilinguisme est l'une des priorités du projet depuis la fin des années 1990. Le français est la deuxième langue. Soixante langues sont représentées en décembre 2010, grâce au patient travail de Distributed Proofreaders, un site web lancé en 2000 pour partager la relecture des livres entre des centaines de volontaires dans de nombreux pays.

\*\*\*

Le Projet Gutenberg est un projet visionnaire lancé en juillet 1971 par Michael Hart pour créer des versions électroniques gratuites d'oeuvres littéraires et les diffuser dans le monde entier. Au 15e siècle, Gutenberg avait permis à chacun d'avoir des livres imprimés pour un prix relativement modique. Au 21e siècle, le Projet Gutenberg permettrait à chacun d'avoir une bibliothèque numérique gratuite.

Michael travaille depuis l'Illinois (États-Unis), dactylographiant des livres du domaine public, par exemple la Bible et les œuvres complètes de Shakespeare, d'abord seul puis avec l'aide de quelques volontaires.

Son projet trouve un nouveau souffle et un rayonnement international avec l'apparition du web en 1990. Comme 95% des usagers de l'internet sont anglophones au début des années 1990, les livres numériques sont surtout en anglais.

Le Projet Gutenberg inspire aussi d'autres bibliothèques numériques en Europe. Le Projekt Runeberg voit le jour en Suède en 1992 dans l'optique de numériser la littérature nordique (scandinave) classique. Le Projekt Gutenberg-DE voit le jour en Allemagne en 1994 dans l'optique de numériser la littérature classique allemande.

Le français fut la deuxième langue du Projet Gutenberg, et c'est toujours le cas en 2010. Les premiers livres disponibles en français sont six oeuvres de Stendhal et deux oeuvres de Jules Verne, toutes mises en ligne début 1997.

Les deux romans de Jules Verne sont «De la terre à la lune» (publié en 1865) et «Le tour du monde en quatre-vingts jours» (publié en 1873). À cette date, les versions anglaises de ces romans sont déjà disponibles en version numérique depuis trois ans, tout comme la version anglaise de «Vingt mille lieues sous les mers» (publié en 1869-1870). Depuis 1994, Jules Verne a toujours fait partie des auteurs les plus téléchargés.

En octobre 1997, Michael Hart annonce son intention d'intensifier la production de livres dans d'autres langues que l'anglais. Début 1998, outre dix titres en français, le catalogue comprend quelques oeuvres en allemand, espagnol, italien et latin. Disponible en mai 1999, l'eBook #2000 est «Don Quijote» (1605) de Cervantès, en espagnol, sa langue originale. En juillet 1999, Michael écrit lors d'un entretien par courriel: «J'introduis une nouvelle langue par mois maintenant, et je vais poursuivre cette politique aussi longtemps que possible.»

Le Project Gutenberg trouve un second souffle avec le lancement de Distributed Proofreaders, un site lancé en octobre 2000 par Charles Franks pour partager la relecture des livres entre des centaines de volontaires dans de nombreux pays.

Disponible en décembre 2000, l'eBook #3000 est le volume 3 (1919) de «À l'ombre des jeunes filles en fleurs» de Marcel Proust, en français.

Disponible en octobre 2001, l'eBook #4000 est «The French Immortals Series» (Recueil de textes d'Immortels français, 1905), en anglais. Cette anthologie regroupe des oeuvres de fiction courtes de plusieurs membres de l'Académie française: Émile Souvestre, Pierre Loti, Hector Malot, Charles de Bernard, Alphonse Daudet et d'autres.

Disponible en avril 2002, l'eBook #5000 est «The Notebooks of Leonardo da Vinci» (Les carnets de Léonard de Vinci), une version anglaise de ces carnets datant du 16<sup>e</sup> siècle. Ce livre est depuis régulièrement présent dans le «Top 100» des livres téléchargés.

On trouve des livres en 25 langues début 2004, en 42 langues en juillet 2005, y compris le sanscrit et les langues mayas, et en 60 langues en novembre 2010. Les dix langues principales sont l'anglais (avec 28.666 ebooks le 6 novembre 2010), le français (1.672 ebooks), l'allemand (715 ebooks), le finnois (542 ebooks), le néerlandais (498 ebooks), le portugais (474 ebooks), le chinois (405 ebooks), l'espagnol (297 ebooks), l'italien (253 ebooks) et le grec (107 ebooks). Ces langues sont suivies du latin, de l'espéranto, du suédois et du tagalog.

Lorsque la traduction automatique aura atteint un taux de fiabilité de 99%, nous pourrons peut-être bénéficier un jour de la traduction immédiate de ces œuvres littéraires dans un vaste choix de langues. Les livres traduits par le biais d'un logiciel de traduction automatique ne rivaliseront certainement pas avec le travail des

traducteurs littéraires et leurs efforts talentueux pendant des jours et des mois sinon des années. Mais ils permettraient au lecteur une première approche d'oeuvres littéraires jamais traduites jusqu'ici, ou traduites seulement dans quelques langues pour des raisons commerciales.

Le texte d'un livre traduit (issu d'un logiciel de traduction) pourrait ensuite être corrigé par des traducteurs (des êtres humains, pas des logiciels), en utilisant une interface semblable à celle utilisée actuellement pour la correction du texte d'un livre (issu d'un logiciel OCR) par les volontaires de Distributed Proofreaders. Nous verrons peut-être un jour un Distributed Translators permettant de partager la révision des traductions, en tant qu'organisme partenaire de Distributed Proofreaders et du Projet Gutenberg.

## **2001 > Wikipédia, encyclopédie collaborative**

### **[Résumé]**

Fondée en janvier 2001 à l'initiative de Jimmy Wales et Larry Sanger (Larry quitte plus tard l'équipe), Wikipédia est une encyclopédie gratuite en ligne écrite collectivement et dont le contenu est librement réutilisable. Sans publicité et financée par des dons, elle est rédigée par des milliers de volontaires, avec possibilité pour tout un chacun d'écrire, corriger ou compléter les articles, aussi bien les siens que ceux d'autres contributeurs. Les articles restent la propriété de leurs auteurs et leur libre utilisation est régie par la licence GFDL (GNU Free Documentation License) ou la licence Creative Commons. En décembre 2006, Wikipédia est l'un des dix sites les plus visités du web, avec 6 millions d'articles dans 250 langues. En 2009, Wikipédia est l'un des cinq sites les plus visités du web, le français étant la troisième langue de l'encyclopédie, après l'anglais et l'allemand. En janvier 2011, Wikipédia fête ses dix ans d'existence avec 17 millions d'articles dans 270 langues et 400 millions de visiteurs par mois pour l'ensemble de ses sites.

\*\*\*

Lancée en janvier 2001, Wikipédia est une encyclopédie gratuite en ligne écrite collectivement et dont le contenu est librement réutilisable.

Qu'est-ce qu'un wiki? Un wiki (terme hawaïen signifiant «vite») est un site web permettant à plusieurs utilisateurs de collaborer simultanément en ligne, en rédigeant le contenu du wiki, en le modifiant et en l'enrichissant en permanence. Le wiki est utilisé par exemple pour créer et gérer des sites d'information, des dictionnaires et des encyclopédies. Le programme présent derrière l'interface d'un wiki est plus ou moins élaboré. Un programme simple gère des textes et des hyperliens. Un programme élaboré permet d'inclure des images, des graphiques, des tableaux, etc.

Fondée à l'initiative de Jimmy Wales et Larry Sanger (Larry Sanger quitte plus tard l'équipe), Wikipédia est immédiatement très populaire. Sans publicité et financée par des dons, elle est rédigée par des milliers de volontaires - qui s'inscrivent sous un pseudonyme - avec possibilité pour tout un chacun d'écrire, corriger et compléter les articles, aussi bien les siens que ceux d'autres contributeurs. Les articles restent la propriété de leurs auteurs et leur libre utilisation est régie par la licence Creative

Commons ou la licence GFDL (GNU Free Documentation License).

Créée en juin 2003, la Wikimedia Foundation gère non seulement Wikipédia mais aussi Wiktionary, un dictionnaire et thésaurus multilingue lancé en décembre 2002, Wikibooks (livres et manuels en cours de rédaction) lancé en juin 2003, auxquels s'ajoutent ensuite Wikiquote (répertoire de citations), Wikisource (textes appartenant au domaine public), Wikimedia Commons (sources multimédia), Wikispecies (répertoire d'espèces animales et végétales), Wikinews (site d'actualités) et enfin Wikiversity (matériel d'enseignement), lancé en août 2006.

En décembre 2004, Wikipédia compte 1,3 million d'articles rédigés par 13.000 contributeurs dans une centaine de langues. En décembre 2006, Wikipédia compte 6 millions d'articles dans 250 langues et devient l'un de dix sites les plus visités du web. En mai 2007, 7 millions d'articles sont disponibles dans 192 langues, dont 1,8 million d'articles en anglais, 589.000 articles en allemand, 500.000 articles en français, 260.000 articles en portugais et 236.000 articles en espagnol. En 2009, l'encyclopédie est l'un des cinq sites les plus visités du web. En septembre 2010, Wikipédia compte 14 millions d'articles dans 272 langues, dont 3,4 millions d'articles en anglais, 1,1 million d'articles en allemand et 1 million d'articles en français, qui est toujours la troisième langue de l'encyclopédie.

Wikipédia fête ses dix ans en janvier 2011 avec 17 millions d'articles dans 270 langues et 400 millions de visiteurs par mois pour l'ensemble de ses sites.

De plus, Wikipédia inspire bien d'autres projets au fil des ans, par exemple Citizendium, lancé en mars 2007 par Larry Sanger en tant qu'encyclopédie collaborative expérimentale au contenu vérifié par des experts, ou encore l'Encyclopedia of Life, un projet global qui voit le jour en mai 2007 pour recenser toutes les espèces animales et végétales connues.

## **2001> L'UNL, projet de métalangage numérique**

### ***[Résumé]***

La Fondation UNDL (Universal Networking Digital Language) est fondée en janvier 2001 à Genève (Suisse) pour développer et promouvoir le projet UNL (Universal Networking Language). Développé à partir de 1996 à Tokyo (Japon) sous l'égide de l'Institute of Advanced Studies (IAS) de l'Université des Nations Unies (UNU), le projet UNL est un projet de métalangage numérique - ou interlangue - formant une passerelle entre une langue source et une langue cible et offrant ainsi une solution aux problèmes de communication posés par la barrière des langues. Comme expliqué en 2010 sur le wiki du projet en langue française, «l'UNL est une langue artificielle créée pour prendre en compte les informations et connaissances véhiculées par les langues humaines. Elle est dotée de composantes lexicales, grammaticales et sémantiques, comme les langues naturelles. Couplée à l'intelligence artificielle, l'UNL facilite la communication entre l'homme et la machine, et par le biais de la machine, entre tous les peuples dans la langue maternelle de chacun.»

\*\*\*

L'UNL (Universal Networking Language) est un projet de métalangage numérique pour l'encodage, le stockage, la recherche et la communication d'informations multilingues.

Il s'agirait d'une interlangue formant une passerelle entre une langue source et une langue cible et offrant ainsi une solution au problème de communication posé par la barrière des langues.

L'UNL est développé à partir de 1996 sous l'égide de l'Institute of Advanced Studies (IAS) de l'Université des Nations Unies (UNU) à Tokyo (Japon) au sein de l'UNL Programme, un programme international impliquant de nombreux partenaires dans plusieurs communautés linguistiques.

En 1998, 120 chercheurs de par le monde travaillent sur un projet plurilingue comportant seize langues (allemand, anglais, arabe, brésilien, chinois, espagnol, français, hindou, indonésien, italien, japonais, letton, mongolien, russe, swahili, thaï).

Au sein de l'IMAG (Institut d'informatique et de mathématiques appliquées de Grenoble), le GETA (Groupe d'étude pour la traduction automatique) participe à l'UNL Programme. Christian Boitet, son directeur, explique en septembre 1998: «Il s'agit non de TAO [traduction assistée par ordinateur] habituelle, mais de communication et recherche d'information multilingue. Quatorze groupes ont commencé le travail sur douze langues (plus deux annexes) depuis début 1997. L'idée est de: (a) développer un standard, dit UNL, qui serait le HTML du contenu linguistique; (b) pour chaque langue, développer un générateur (dit "déconvertisseur") accessible sur un ou plusieurs serveurs, et un "enconvertisseur".»

Les applications possibles sont le courriel multilingue, les informations multilingues, les dictionnaires numériques pour la lecture des langues étrangères sur le web et enfin la traduction automatique pour la navigation sur le web et la veille informatique.

Quelles sont les perspectives? D'après Christian Boitet, «le plan général est d'ouvrir le projet aux autres langues de l'ONU en 2000. Il faudrait arriver à un état satisfaisant pour les douze autres avant. Du point de vue politique et culturel, ce projet est très important, en ce qu'il montre pour la première fois une voie possible pour construire divers outils soutenant l'usage de toutes les langues sur internet, qu'elles soient majoritaires ou minoritaires. En particulier, ce devrait être un projet majeur pour la Francophonie.»

Ce programme se poursuit ensuite sous l'égide de la Fondation UNDL (Universal Networking Digital Language), créée en janvier 2001 à Genève (Suisse) pour développer et promouvoir le projet UNL, en partenariat avec les Nations Unies.

Comme expliqué en 2010 sur le wiki en langue française du projet, «l'UNL est une langue artificielle créée pour prendre en compte les informations et connaissances véhiculées par les langues humaines. Elle est dotée de composantes lexicales, grammaticales et sémantiques, comme les langues naturelles. Couplée à

l'intelligence artificielle, l'UNL facilite la communication entre l'homme et la machine, et par le biais de la machine, entre tous les peuples dans la langue maternelle de chacun. Notre première tâche est de compléter le système UNL. Ensuite, le mettre au service des toutes les nations.»

## **2001 > Un marché pour les logiciels de traduction**

### **[Résumé]**

En mars 2001, IBM se lance dans un marché de la traduction en pleine expansion avec un produit professionnel haut de gamme, le WebSphere Translation Server. Ce logiciel traduit instantanément en plusieurs langues (allemand, anglais, chinois, coréen, espagnol, français, italien, japonais) les pages web, courriels et chats. Il interprète 500 mots à seconde et permet l'ajout de termes spécifiques. Par ailleurs, des logiciels de traduction assistée par ordinateur (TAO), destinés aux traducteurs professionnels, incluent une «mémoire de traduction» avec gestion de la terminologie en temps réel et contrôle typographique, par exemple Wordfast, créé dès 1999 par Yves Champollion et compatible avec les autres grands logiciels du marché que sont le WebSphere Translation Server d'IBM et les logiciels de SDL Trados. Utilisable sur toute plateforme (Windows, Mac, Linux), Wordfast compte 14.000 clients dans le monde en 2010, dont les Nations Unies, Coca-Cola et Sony.

\*\*\*

L'internationalisation de l'internet favorise un marché pour les logiciels de traduction, tout comme le développement du commerce électronique.

Les sociétés Systran, Alis Technologies, Lernout & Hauspie, Globalink, Softissimo et bien d'autres développent des logiciels, produits et services ciblant trois types de clients: le grand public, les professionnels des langues et les sociétés localisant leurs sites web.

En mars 2001, IBM se lance dans un marché en pleine expansion avec un produit professionnel haut de gamme, le WebSphere Translation Server. Ce logiciel traduit instantanément en plusieurs langues (allemand, anglais, chinois, coréen, espagnol, français, italien, japonais) les pages web, courriels et chats. Il interprète 500 mots à la seconde et permet l'ajout de vocabulaires spécifiques.

Un logiciel de traduction automatique (TA) analyse le texte dans la langue à traduire (langue source) et génère automatiquement le texte dans la langue désirée (langue cible), en utilisant des règles précises pour le transfert de la structure grammaticale. L'être humain n'intervient pas au cours du processus, contrairement à la traduction assistée par ordinateur (TAO), qui implique une interaction entre l'homme et la machine.

Par ailleurs, des logiciels de traduction assistée par ordinateur (TAO) sont proposés aux traducteurs professionnels. Lancé en 1999 à Paris par Yves Champollion, Wordfast est un logiciel incluant une «mémoire de traduction» avec gestion de la terminologie en temps réel et contrôle typographique. Il est compatible avec les

autres grands logiciels du marché que sont le WebSphere Translation Server d'IBM et les logiciels de SDL Trados. Utilisable sur toute plateforme (Windows, Mac, Linux), Wordfast compte 14.000 clients dans le monde en 2010, dont les Nations Unies, Coca-Cola, Sony et bien d'autres.

Selon Tim McKenna, écrivain et philosophe, interviewé en octobre 2000, «lorsque la qualité des logiciels sera suffisante pour que les gens puissent converser par écrit et par oral sur le web en temps réel dans différentes langues, nous verrons tout un monde s'ouvrir à nous. Les scientifiques, les hommes politiques, les hommes d'affaires et bien d'autres groupes seront à même de communiquer immédiatement entre eux sans l'intermédiaire de médiateurs ou traducteurs.»

Selon Randy Hobler, consultant en marketing internet de produits et services de traduction, l'étape suivante sera celle de la «transparence transculturelle et transnationale».

Il explique dès septembre 1998: «Nous arriverons rapidement au point où une traduction très fidèle du texte et de la parole sera si commune qu'elle pourra faire partie des plateformes ou même des puces. À ce stade, lorsque le développement de l'internet aura atteint sa vitesse de croisière, lorsque la fidélité de la traduction atteindra plus de 98% et lorsque les différentes combinaisons de langues possibles auront couvert la grande majorité du marché, la transparence de la langue - à savoir toute communication d'une langue à une autre - sera une vision trop restrictive pour ceux qui vendent cette technologie. Le développement suivant sera la "transparence transculturelle et transnationale" dans laquelle les autres aspects de la communication humaine, du commerce et des transactions au-delà du seul langage entreront en scène. Par exemple, les gestes ont un sens, les mouvements faciaux ont un sens, et ceci varie en fonction des normes sociales d'un pays à l'autre. (...)

Les cultures diffèrent de milliers de façons, et la plupart de leurs codes peuvent être modifiés par voie informatique lorsqu'on passe d'un code culturel à l'autre. Ceci inclut les lois, les coutumes, les habitudes de travail, l'éthique, le change monétaire, les différences de taille dans les vêtements, les différences entre le système métrique et le système de mesure anglophone, etc. Les firmes dynamiques répertorieront et programmeront ces différences, et elles vendront des produits et services afin d'aider les habitants de la planète à mieux communiquer entre eux. Une fois que ces produits et services seront largement répandus, ils contribueront réellement à une meilleure compréhension à l'échelle internationale.»

## **2004 > Le web 2.0, communauté et partage**

### ***[Résumé]***

Le terme «web 2.0» émane d'un éditeur de livres informatiques, Tim O'Reilly, qui l'utilise pour la première fois en 2004 en tant que titre d'une série de conférences qu'il est en train d'organiser. Le web 2.0 est caractérisé par les notions de communauté et de partage, avec une flopée de sites dont le contenu est alimenté par les utilisateurs, par exemple les blogs, les wikis, les sites sociaux et les

encyclopédies collaboratives. Wikipédia, Facebook et Twitter bien sûr, mais aussi des dizaines de milliers d'autres. Le web 2.0 tente de répondre au rêve formulé par Tim Berners-Lee, inventeur du web en 1990, qui écrit dans un essai daté d'avril 1998: «Le rêve derrière le web est un espace d'information commun dans lequel nous communiquons en partageant l'information. Son universalité est essentielle, à savoir le fait qu'un lien hypertexte puisse pointer sur quoi que ce soit, quelque chose de personnel, de local ou de global, aussi bien une ébauche qu'une réalisation très sophistiquée.»

\*\*\*

Le terme «web 2.0» émane d'un éditeur de livres informatiques, Tim O'Reilly, qui l'utilise pour la première fois en 2004 en tant que titre d'une série de conférences qu'il est en train d'organiser.

Le web 2.0 est caractérisé par les notions de communauté et de partage, avec une flopée de sites dont le contenu est alimenté par les utilisateurs, par exemple les blogs, les wikis, les sites sociaux et les encyclopédies collaboratives. Wikipédia, Facebook et Twitter bien sûr, mais aussi des dizaines de milliers d'autres.

### ***Les blogs envahissent la toile***

Un blog (ou blogue) est un journal en ligne tenu par une personne ou un groupe. Ce journal est le plus souvent présenté par ordre chronologique inversé (du plus récent au plus ancien) et il est actualisé d'heure en heure ou bien une fois par mois. Le premier blog apparaît en 1997. En 2004, Le Monde.fr, site du quotidien Le Monde, lance ses propres blogs, «un formidable format d'expression journalistique qui permet un dialogue quasi-instantané avec son lecteur», selon Yann Chapellon, directeur du Monde interactif. En juillet 2005, il y aurait 14 millions de blogs dans le monde, avec 80.000 nouveaux blogs par jour. En décembre 2006, Technorati, moteur de recherche pour blogs puis site spécialisé, recense 65 millions de blogs, avec 175.000 nouveaux blogs par jour. Certains blogs sont consacrés aux photos (photoblogs), à la musique (audioblogs ou podcasts) et aux vidéos (vidéoblogs ou vlogs).

### ***Les wikis, sites collaboratifs***

Un wiki (terme hawaïen signifiant «vite») est un site web permettant à plusieurs utilisateurs de collaborer en ligne sur un même projet. Le concept du wiki devient très populaire en 2000, avec possibilité pour les participants de contribuer à la rédaction du contenu, de modifier ce contenu et de l'enrichir en permanence. Le wiki est utilisé par exemple pour créer et gérer des sites d'information, des dictionnaires et des encyclopédies. Le programme présent derrière l'interface d'un wiki est plus ou moins élaboré. Un programme simple gère des textes et des hyperliens. Un programme élaboré permet d'inclure des images, des graphiques, des tableaux, etc. L'encyclopédie wiki la plus connue est Wikipédia.

### ***Facebook, réseau social***

Facebook est un réseau social fondé en février 2004 par Mark Zuckerberg et ses collègues étudiants. Destiné à l'origine aux étudiants de l'Université de Harvard, puis aux étudiants de toutes les universités américaines, le réseau social s'ouvre au monde en septembre 2006 afin de connecter entre eux des personnes proches (famille, amis, collègues) ou des personnes partageant les mêmes centres d'intérêt. En juin 2010, Facebook devient le deuxième site mondial en nombre de visites, après Google, et fête ses 500 millions d'utilisateurs tout en suscitant des débats sur le respect de la vie privée.

### ***Twitter, l'information en 140 caractères***

Lancé en 2006 par Jack Dorsey, Twitter est un outil de réseau social et de micro-blogging permettant à l'utilisateur d'envoyer gratuitement des tweets (messages brefs au format texte) de 140 caractères maximum, par messagerie instantanée, par SMS ou via l'internet. Parfois décrit comme le SMS de l'internet, Twitter gagne rapidement une popularité mondiale, avec 106 millions d'utilisateurs en avril 2010 et 300.000 nouveaux usagers par jour. Quant aux tweets, on compte 5.000 tweets quotidiens en 2007, 300.000 en 2008, 2,5 millions en 2009, 50 millions en janvier 2010 et 55 millions en avril 2010, avec un archivage systématique des tweets à usage public par la Bibliothèque du Congrès en tant que reflet des tendances de notre époque.

### ***Le rêve de Tim Berners-Lee***

Comme on le voit, le web 2.0 tente de répondre au rêve formulé par Tim Berners-Lee, inventeur du web en 1990, qui écrit dans un essai daté d'avril 1998: «Le rêve derrière le web est un espace d'information commun dans lequel nous communiquons en partageant l'information. Son universalité est essentielle, à savoir le fait qu'un lien hypertexte puisse pointer sur quoi que ce soit, quelque chose de personnel, de local ou de global, aussi bien une ébauche qu'une réalisation très sophistiquée. Deuxième partie de ce rêve, le web deviendrait d'une utilisation tellement courante qu'il serait un miroir réaliste (sinon la principale incarnation) de la manière dont nous travaillons, jouons et nouons des relations sociales. Une fois que ces interactions seraient en ligne, nous pourrions utiliser nos ordinateurs pour nous aider à les analyser, donner un sens à ce que nous faisons, et voir comment chacun trouve sa place et comment nous pouvons mieux travailler ensemble.» (extrait de «The World Wide Web: A very short personal history»)

## **2007 > La norme ISO 639-3 pour identifier les langues**

### ***[Résumé]***

Le premier standard est la norme ISO 639-1, adoptée par l'Organisation internationale de normalisation (ISO) en 1988 et qui identifie chaque langue sur deux lettres. Suit la norme ISO 639-2, publiée en 1998 pour identifier 400 langues sur trois lettres. Par ailleurs, l'Ethnologue, catalogue encyclopédique de langues vivantes publié par SIL International, développe également ses propres codes sur trois lettres dans sa base de données depuis 1971, avec inclusion de ces identifiants dans l'encyclopédie depuis 1984 (10e édition). En 2002, à l'invitation de l'Organisation

internationale de normalisation, SIL International prépare une nouvelle norme ISO qui harmonise les identifiants utilisés dans l'Ethnologue avec ceux de la norme ISO 639-2, en intégrant aussi les identifiants des langues mortes et artificielles utilisés dans la Linguist List, une grande liste de diffusion à destination des linguistes. Publiée en 2007, la norme ISO 639-3 attribue un identifiant de trois lettres à 7.589 langues. SIL International est également désigné comme l'organisme responsable de la gestion du cycle annuel des modifications et des mises à jour.

\*\*\*

Publiée en 2007, la norme ISO 639-3 attribue un identifiant de trois lettres à 7 589 langues, par exemple « fra » pour le français.

Cette norme existe depuis plus de trente ans. La première norme en vigueur est la norme ISO 639-1, adoptée par l'Organisation internationale de normalisation (ISO) en 1988 et qui identifie chaque langue sur deux lettres.

Dix ans plus tard suit la norme ISO 639-2, adoptée en 1998 et qui identifie cette fois chaque langue au moyen de trois lettres mais se limite à 400 langues. Cette norme est la convergence de la norme ISO 639-1 avec la norme ANSI Z39.53 (ANSI : American National Standards Institute). La norme ANSI correspond aux codes de langues MARC (Machine Readable Cataloging), des identifiants sur trois lettres développés par les bibliothèques américaines et adoptés en tant que norme nationale en 1987.

Par ailleurs, un autre effort de codification est poursuivi par l'Ethnologue, magnifique catalogue encyclopédique de langues vivantes publié depuis 1950 par SIL International, avec une nouvelle version tous les quatre ans. L'équipe de l'Ethnologue a elle aussi créé des codes de trois lettres pour chaque langue dans sa base de données, ce depuis 1971, avec inclusion des codes dans l'encyclopédie elle-même à partir de la 10<sup>e</sup> édition (1984).

Revenons à la norme ISO 639-2, qui devient vite insuffisante du fait du petit nombre de langues pris en compte. En 2002, l'ISO invite donc SIL International à établir une nouvelle norme qui harmoniserait les identifiants utilisés par l'Ethnologue avec les identifiants de la norme ISO 639-2, en y intégrant aussi les codes des langues mortes utilisés par la Linguist List, grande liste de diffusion à destination des linguistes.

Le résultat, officiellement approuvé en 2006 et publié en 2007, est la norme ISO 639-3, qui attribue un code de trois lettres à 7 589 langues, avec un travail aussi exhaustif que possible puisque cette liste de codes inclut les langues vivantes ou mortes, anciennes ou artificielles, importantes ou minoritaires, écrites ou non écrites.

SIL International est également désigné comme l'organisme responsable de l'enregistrement de nouvelles langues pour l'ISO 639-3, et gère donc le cycle annuel des modifications et des mises à jour. Au 21<sup>e</sup> siècle, la préservation des langues passe par leur codification.

## 2007 > Google Traduction

### **[Résumé]**

Lancé par Google en octobre 2007, Google Traduction est un service en ligne gratuit qui traduit instantanément un texte ou une page web dans une autre langue. Les usagers copient un texte dans l'interface web ou entrent une adresse web. Le service de traduction automatique de Google se base sur une analyse statistique et non sur une analyse traditionnelle basée sur des règles. Avant cette date, Google utilisait un traducteur de Systran du même type que Babel Fish dans Yahoo! Comme tout outil de traduction automatique, Google Traduction peut aider l'utilisateur à comprendre le sens général d'un texte en langue étrangère, mais ne propose pas de traductions exactes. En 2009, le texte peut être lu par synthèse vocale, avec l'ajout de nouvelles langues au fil des mois. Disponible en juin 2009, Google Translator Toolkit est un service web permettant aux traducteurs de réviser les traductions générées automatiquement par Google Traduction.

\*\*\*

Lancé en octobre 2007, Google Traduction (Google Translate) est un service en ligne gratuit qui traduit instantanément un texte ou une page web dans une autre langue. Les usagers copient un texte dans l'interface web ou entrent une adresse web.

Innovation par rapport à la concurrence, Google Traduction se base sur une analyse statistique pour la traduction automatique et non sur une analyse traditionnelle basée sur des règles.

Google s'empresse lui aussi d'expliquer que ce nouveau service peut aider l'utilisateur à comprendre le sens général d'un texte en langue étrangère, mais ne propose pas de traductions exactes. (En janvier 2011, les usagers ont le choix entre plusieurs traductions pour les mêmes mots.)

Avant le lancement de Google Translate, Google utilisait un traducteur de Systran du même type que Babel Fish, avec plusieurs étapes quant aux paires de langues disponibles.

Étape un: de l'anglais vers le français, l'allemand et l'espagnol, et vice versa.

Étape deux: de l'anglais vers le portugais et le flamand, et vice versa.

Étape trois: de l'anglais vers l'italien, et vice versa.

Étape quatre: de l'anglais vers le chinois simplifié, le japonais et le coréen, et vice versa.

Étape cinq (avril 2006): de l'anglais vers l'arabe, et vice versa.

Étape six (décembre 2006): de l'anglais vers le russe, et vice versa.

Étape sept (février 2007): de l'anglais vers le chinois traditionnel, et du chinois simplifié vers le chinois traditionnel, et vice versa.

Voici les étapes propres au système de traduction de Google, tout au moins les dix premières.

Étape un (octobre 2007): toutes les langues disponibles jusqu'ici, avec toutes les combinaisons possibles.

Étape deux: de l'anglais à l'hindou, et vice versa.

Étape trois (mai 2008): bulgare, croate, danois, finlandais, grec, néerlandais, norvégien, polonais, roumain, suédois, tchèque, avec toutes les combinaisons possibles.

Étape quatre (septembre 2008): catalan, hébreu, indonésien, letton, lituanien, philippin, serbe, slovaque, slovène, ukrainien, vietnamien.

Étape cinq (janvier 2009): albanais, estonien, galicien, hongrois, maltais, thaï, turc.

Étape six (juin 2009): perse.

Étape sept (août 2009): afrikaans, biélorusse, gallois, irlandais, islandais, macédonien, malais, swahili, yiddish.

Étape huit (janvier 2010): haïtien créole.

Étape neuf (mai 2010): arménien, azéri, basque, géorgien, ourdou.

Étape dix (octobre 2010): latin.

Etc.

En 2009, nouvelle innovation, le texte peut être lu par synthèse vocale, avec l'ajout de nouvelles langues au fil des mois pour la synthèse vocale.

La même année, deuxième innovation, Google lance le Google Translator Toolkit, un service web permettant aux traducteurs (humains) de réviser les traductions générées automatiquement par Google Traduction. Ils peuvent aussi partager traductions, glossaires et mémoires de traduction.

## **2009 > 6.909 langues vivantes dans l'Ethnologue**

### **[Résumé]**

6.909 langues vivantes sont répertoriées dans la 16e édition (2009) de l'Ethnologue. Ce catalogue encyclopédique comprend deux versions: une version web gratuite depuis 1996 et une version imprimée payante depuis 1950. Une version CD-Rom payante est également disponible dans les années 1990 et abandonnée ensuite. Publiée par SIL International (SIL: Summer Institute of Linguistics), cette oeuvre de référence répertorie les langues selon divers critères (nom de la langue, famille linguistique, pays dans lequel la langue est parlée, identifiant de trois lettres, etc.) tout en offrant un moteur de recherche unique, des index et des cartes géographiques. Un petit groupe de chercheurs travaillant à Dallas, dans le Texas, coordonne le travail de milliers de linguistes qui glanent et vérifient des informations dans le monde entier. Une nouvelle version de l'Ethnologue est publiée tous les quatre ans environ.

\*\*\*

6.909 langues vivantes sont répertoriées dans la 16e édition (2009) de l'Ethnologue, un catalogue encyclopédique comprenant deux versions: une version web gratuite depuis 1996 et une version imprimée payante depuis 1950.

Une version CD-Rom payante est également disponible dans les années 1990 et abandonnée ensuite.

Publié par SIL International (SIL: Summer Institute of Linguistics), cette oeuvre de référence, dont le titre complet est «The Ethnologue: Languages of the World», répertorie les langues selon divers critères (nom de la langue, famille linguistique, pays dans lequel la langue est parlée, identifiant de trois lettres, etc.) tout en offrant un moteur de recherche unique, des index et des cartes géographiques. Une nouvelle version de l'Ethnologue est publiée tous les quatre ans environ.

Débuté en 1950 pour offrir un catalogue des langues minoritaires avant de s'élargir à toutes les langues vivantes de la planète, ce travail est mené sous l'égide d'une petite équipe de chercheurs basée à Dallas, dans le Texas. Cette équipe rassemble et organise la masse d'informations glanées et vérifiées une à une sur le terrain par des milliers de linguistes regroupés en équipes nationales et/ou linguistiques présentes sur tous les continents.

Barbara Grimes, directrice de publication entre 1971 et 2000 (8e-14e éditions), relate en janvier 2000: «Il s'agit d'un catalogue des langues dans le monde, avec des informations sur les pays où elles sont parlées, une estimation du nombre de personnes qui les parlent, la famille linguistique à laquelle elles appartiennent, les autres termes utilisés pour ces langues, les noms de dialectes, diverses informations socio-linguistiques et démographiques, les dates des Bibles publiées, un index des noms de langues [Ethnologue Name Index], un index des familles linguistiques [Ethnologue Language Family Index] et enfin des cartes géographiques pour les langues.»

Mais qu'est-ce exactement qu'une langue? Dans l'introduction de la 16e édition (2009) de l'Ethnologue, on peut lire ceci: «La manière dont chacun choisit de définir une langue dépend des motifs qu'on a d'identifier cette langue comme étant distincte d'une autre. Certains basent la définition d'une langue sur des raisons purement linguistiques. D'autres reconnaissent la nécessité de prendre également en compte des facteurs sociaux, culturels ou politiques. En outre, les locuteurs d'une langue ont souvent leurs propres critères sur l'appropriation d'une langue comme étant la leur. Ces critères sont souvent bien davantage liés à des questions de patrimoine et d'identité qu'aux traits linguistiques de la langue ou des langues en question.»

Comme expliqué dans cette même introduction, une caractéristique de la base de données de l'Ethnologue depuis 1971 est un système de codes qui identifient chaque langue sur trois lettres (par exemple «fra» pour le français), avec inclusion des identifiants dans l'encyclopédie elle-même à partir de la 10e édition (1984). À l'invitation de l'Organisation internationale de normalisation (ISO) en 2002, SIL International prépare une nouvelle norme ISO permettant d'harmoniser les identifiants utilisés dans l'Ethnologue avec ceux de la norme ISO 639-2 (1998), en intégrant aussi les identifiants des langues mortes et artificielles utilisés dans la Linguist List, une grande liste de diffusion à destination des linguistes. Publiée en 2007, la nouvelle norme ISO 639-3 attribue un identifiant de trois lettres à près de 7.500 langues. SIL International est également désigné comme l'organisme responsable de la gestion du cycle annuel des modifications et des mises à jour.

## 2010 > Un atlas de l'UNESCO pour les langues menacées

### **[Résumé]**

En 2010, dans le cadre de son programme de préservation des langues menacées, l'UNESCO (Organisation des Nations Unies pour l'éducation, la science et la culture) lance un atlas interactif des langues en danger dans le monde. La version en ligne gratuite est complémentaire de la version imprimée payante (3<sup>e</sup> édition, 2010), réalisée sous la direction de Christopher Moseley, et disponible en anglais, en français et en espagnol, suite aux deux premières éditions publiées en 1996 et 2001. L'atlas interactif comprend 2.473 langues en juin 2011, avec un moteur de recherche par pays ou région, par nom de langue, par nombre de locuteurs, par vitalité et par code ISO 639-3. Les noms des langues sont indiqués dans leurs transcriptions en français, en anglais et en espagnol. Les noms alternatifs (variantes orthographiques, dialectes ou noms en caractères non latins) sont également fournis dans de nombreux cas.

\*\*\*

En 2010, dans le cadre de son programme de préservation des langues menacées, l'UNESCO (Organisation des Nations Unies pour l'éducation, la science et la culture) lance un atlas interactif des langues en danger dans le monde.

La version en ligne gratuite est complémentaire de la version imprimée payante, la dernière édition en date étant la 3<sup>e</sup> édition (2010), réalisée sous la direction de Christopher Moseley et publiée en trois langues (français, anglais, espagnol). Les deux premières éditions datent respectivement de 1996 et 2001.

L'atlas interactif comprend 2.473 langues en juin 2011, avec un moteur de recherche par pays ou région, par nom de langue, par nombre de locuteurs de/à, par vitalité et par code ISO 639-3.

Les noms des langues sont indiqués dans leurs transcriptions en français, en anglais et en espagnol. Les noms alternatifs (variantes orthographiques, dialectes ou noms en caractères non latins) sont également fournis dans de nombreux cas.

### ***La vitalité des langues***

Le rapport de l'UNESCO sur la vitalité et le danger de disparition des langues établit six niveaux de vitalité: sûre, vulnérable, en danger, sérieusement en danger, en situation critique, éteinte.

«Sûre» signifie que la langue est parlée par toutes les générations et que la transmission intergénérationnelle est ininterrompue. Les langues concernées ne sont donc pas incluses dans l'atlas.

«Vulnérable» signifie que la plupart des enfants parlent la langue, mais qu'elle est restreinte à certains domaines, par exemple la maison.

«En danger» signifie que les enfants n'apprennent plus la langue comme langue maternelle à la maison.

«Sérieusement en danger» signifie que la langue est parlée par les grands-parents. Si la génération des parents peut la comprendre, les parents ne la parlent pas entre eux ou avec les enfants.

«En situation critique» signifie que les locuteurs les plus jeunes sont les grands-parents et leurs ascendants, et qu'ils ne parlent la langue que partiellement et peu fréquemment.

«Éteinte» signifie qu'il n'y a plus de locuteurs. L'atlas inclut les langues éteintes depuis les années 1950.

### ***Comment définir une langue en péril***

À quel moment une langue est-elle considérée comme en péril? Comme expliqué par l'UNESCO sur le site de l'atlas interactif, «une langue est en péril lorsque ses locuteurs cessent de l'utiliser, réservent son usage à des domaines de plus en plus restreints, emploient un moins grand nombre de registres ou de styles de parole, et/ou arrêtent de la transmettre à la génération suivante. Aucun facteur ne détermine à lui seul si une langue est en danger.»

Selon les experts de l'UNESCO, il importe de considérer les neuf critères suivants: (1) la transmission de la langue d'une génération à l'autre, (2) le nombre absolu de locuteurs, (3) le taux de locuteurs par rapport à l'ensemble de la population, (4) l'utilisation de la langue dans les différents domaines publics et privés, (5) la réactivité d'une langue face aux nouveaux domaines et médias, (6) l'existence de matériels d'apprentissage et d'enseignement de la langue, (7) les attitudes et politiques linguistiques au niveau du gouvernement et des institutions, y compris son usage et son statut au niveau officiel, (8) les attitudes des membres de la communauté vis-à-vis de leur propre langue, (9) le type et la qualité de la documentation.

Quels sont les facteurs de disparition d'une langue? Selon les mêmes experts, «une langue disparaît lorsqu'elle n'a plus de locuteurs ou que ceux-ci se mettent à parler une autre langue - en général, une langue de plus grande importance utilisée par un groupe plus puissant. Les langues sont menacées par des forces externes telles qu'une domination militaire, économique, religieuse, culturelle ou éducative, ou par des forces internes comme l'attitude négative d'une population à l'égard de sa propre langue. Aujourd'hui, les migrations accrues et l'urbanisation rapide s'accompagnent souvent de la perte des modes de vie traditionnels et d'une forte pression en faveur de l'utilisation d'une langue dominante qui est nécessaire - ou perçue comme telle - à une vraie participation totale à la vie civique et au progrès économique.»