

This is a \*copyrighted\* Project Gutenberg eBook, details below.

Title: Le multilinguisme sur le Web

Author: Marie Lebert

Release date: October 26, 2008 [eBook #27029]

Language: French

\*\*\* START OF THE PROJECT GUTENBERG EBOOK LE MULTILINGUISME SUR LE WEB \*\*\*

Produced by Al Haines

# LE MULTILINGUISME SUR LE WEB

**MARIE LEBERT**

CEVEIL, Montréal, 1999 & NEF, University of Toronto, 2001

Copyright © 1999 Marie Lebert

Datée de février 1999, cette étude s'articule autour de quatre grands thèmes: le multilinguisme, les ressources linguistiques, la traduction et la recherche. Elle est basée sur de nombreux entretiens. Elle est complétée par une sélection de sites. Le titre de la version anglaise est: Multilingualism on the Web. Les versions originales sont disponibles sur le NEF: <http://www.etudes-francaises.net/entretiens/multi.htm>

## TABLE

1. Introduction
2. Le multilinguisme
3. Ressources linguistiques
4. Traduction
5. Recherche
6. Sélection de sites
7. Index des sites et pages web

## 1. INTRODUCTION

On déclare souvent qu'Internet abolit le temps, les distances et les frontières, mais qu'en est-il des langues?

Dès ses débuts, la principale langue d'Internet a été l'anglais. Même si l'anglais est toujours prédominant en 1998-99, d'autres langues sont maintenant présentes, et leur progression est constante. Il reste à souhaiter que, tôt ou tard, la répartition des langues sur Internet corresponde à la répartition des langues dans le monde, et que des logiciels de traduction soient disponibles gratuitement sur le Web pour traduire instantanément une page dont on ne comprend pas la langue. Ceci dit, il reste beaucoup à faire pour que ce rêve devienne réalité.

Cette étude s'articule autour de quatre grands thèmes: le multilinguisme, les ressources linguistiques, la traduction et la recherche.

Dans le chapitre concernant le multilinguisme, on étudiera le développement des langues non anglophones sur le Web, en prenant le français à titre d'exemple. On évoquera aussi la diversité des langues dans l'Union européenne.

Dans le chapitre relatif aux ressources linguistiques, on donnera quelques exemples des ressources disponibles sur le Web: sites répertoriant les ressources linguistiques, répertoires de langues, dictionnaires et glossaires, bases textuelles et terminologiques.

Dans le chapitre ayant trait à la traduction, on explorera les perspectives apportées par la traduction automatique et la traduction assistée par ordinateur.

Dans le dernier chapitre consacré à la recherche, on présentera plusieurs projets de recherche dans les domaines suivants: traduction automatique, linguistique computationnelle, ingénierie du langage, internationalisation et localisation.

En août puis en décembre 1998, j'ai envoyé une enquête basée sur trois questions à un certain nombre d'organisations et de sociétés dont j'avais découvert les sites sur le Web. Les trois questions étaient les suivantes:

a) Comment voyez-vous le multilinguisme sur le Web?

b) Quel a été le bénéfice de l'utilisation d'Internet dans votre vie professionnelle ou dans l'activité de votre organisme ou société?

c) Comment voyez-vous l'évolution vers un Internet multilingue, dans le cadre de votre avenir professionnel ou en général?

Les réponses reçues sont incluses dans les pages qui suivent. Mes remerciements les plus vifs vont à tous ceux qui ont contribué à cette étude.

En tant que traductrice-éditrice (travaillant essentiellement pour le Bureau international du Travail (BIT), situé à Genève), je suis fascinée par les langues, et je souhaitais donc en savoir plus sur le multilinguisme sur le Web. Durant le deuxième semestre 1998, j'ai passé un certain nombre d'heures à explorer le Web dans cette intention. Datée de février 1999, cette étude est le résultat de ces recherches. Bien que très proche de la version anglaise, elle n'est pas la traduction littérale de celle-ci. Cette étude ne cherche pas non plus à être exhaustive. Elle vise seulement à diffuser mes propres recherches dans les domaines qui m'intéressent particulièrement. Je m'intéresse également aux relations entre l'imprimé et Internet, objet d'une autre étude.

## 2. LE MULTILINGUISME

[Dans ce chapitre:]

[2.1. Le Web: d'abord anglais, ensuite multilingue / 2.2. Le français face au géant anglais / 2.3. Diversité des langues dans l'Union européenne]

2.1. Le Web: d'abord anglais, ensuite multilingue

A ses débuts, Internet était anglophone à pratiquement 100%, ce qui s'explique par le fait qu'il a été créé aux Etats-Unis en tant que réseau mis en place dès 1969 par le Pentagone avant de se développer dans les organismes gouvernementaux et les universités.

Après la création du World Wide Web en 1989-90 par Tim Berners-Lee au Laboratoire européen pour la physique des particules (CERN) (Genève, Suisse) et la diffusion en novembre 1993 du premier navigateur (Mosaic, ancêtre de Netscape), Internet s'est développé de manière foudroyante, d'abord aux Etats-Unis grâce aux investissements considérables réalisés par le gouvernement, puis au Canada et dans le reste du monde.

En 1997, Babel, une initiative conjointe d'Alis Technologies et de l'Internet Society, a mené la première étude d'ensemble sur la répartition réelle des langues sur Internet. Les résultats sont publiés dans le Palmarès des langues de la Toile, daté de juin 1997 et qui devrait être réactualisé prochainement. Les pourcentages étaient 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.

Le fait que l'anglais soit toujours représenté par un très fort pourcentage s'explique par les trois facteurs suivants: 1) l'anglais est la principale langue d'échange internationale, 2) les premières années ont vu la création d'un grand nombre de sites web émanant des Etats-Unis, du Canada ou du Royaume-Uni, 3) la proportion de cybernautes est encore particulièrement forte en Amérique du Nord par rapport au reste du monde, pour les raisons suivantes: a) jusque très récemment, le matériel informatique et les logiciels étaient bien meilleur marché qu'ailleurs, b) on ne paie qu'un forfait mensuel pour les communications téléphoniques locales, ce qui rend l'utilisation d'Internet très économique par rapport à l'Europe, où les communications locales sont tarifées à la durée.

Dans La Francophonie en quête d'identité sur le Web, un article de Hugues Henry publié par le cyberquotidien Multimédium, Jean-Pierre Cloutier, auteur des Chroniques de Cybérie, expliquait:

"Au Québec, je passe environ 120 heures par mois en ligne. Mon accès Internet me coûte 30 \$ [canadiens], si j'ajoute mon forfait téléphonique d'environ 40 \$ (avec les divers services facultatifs dont je dispose), le coût total de mon branchement s'établit donc à 70 \$ par mois. Je vous laisse deviner ce qu'il m'en coûterait en France, en Belgique ou en Suisse, où les communications locales sont facturées à la minute, pour le même nombre d'heures en ligne."

Il s'ensuit que de nombreux cybernautes européens passent beaucoup moins de temps sur le Web qu'ils ne le souhaiteraient, ou choisissent de "surfer" la nuit pour éviter les factures trop élevées. Fin 1998 et début 1999, des mouvements de grève ont eu lieu dans plusieurs pays d'Europe pour faire pression sur les sociétés prestataires de services téléphoniques afin qu'elles baissent leur prix et qu'elles proposent des forfaits.

Dans un article daté du 21 juillet 1998 et publié dans ZDNN (ZD Network News), Martha L. Stone précisait:

"Cette année, le nombre de nouveaux sites non anglophones va probablement dépasser celui de nouveaux sites anglophones, et le monde cyber est en train de véritablement devenir une toile à l'échelle mondiale. [...] Selon Global Reach, les groupes de nouveaux sites web qui se développent le plus vite sont les groupes non anglophones: on note une progression de 22,4% pour les sites espagnols, 12,3% pour les sites japonais, 14% pour les sites allemands et 10% pour les sites francophones. On estime à 55,7 millions le nombre de personnes non anglophones ayant accès au Web. [...] Alors que 6% seulement de la population mondiale est de langue maternelle anglaise (et 16% d'hispanophones), 80% des pages web sont en anglais."

En contrepartie, même si l'anglais est encore prédominant, il faut cependant noter que de plus en plus de sociétés et d'organismes anglophones prennent conscience du fait que tout le monde ne comprend pas l'anglais. Les sites bilingues ou multilingues sont de plus en plus nombreux, ce pour des raisons aussi bien commerciales que culturelles.

Sensible à ce problème, AltaVista, moteur de recherche utilisé par douze millions d'internautes, mettait en place en décembre 1997 AltaVista Translation, un service de traduction automatisée de l'anglais vers cinq autres langues (allemand, espagnol, français, italien et portugais), et vice versa. Alimenté par des dictionnaires multilingues contenant plus de 2,5 millions de termes, ce service, gratuit et instantané, a été créé par Systran, société franco-américaine pionnière dans le domaine de la traduction automatique.

La traduction étant entièrement automatisée, elle est évidemment approximative. Le texte à traduire doit être de trois pages maximum, et la page originale et la traduction apparaissent en vis-à-vis sur l'écran. Cet outil a évidemment ses limites, mais il a le mérite d'exister et il préfigure ceux de demain.

De par sa vocation internationale, Internet doit être multilingue. On dispose enfin d'un instrument qui peut abolir les frontières au lieu d'en créer d'autres. De plus en plus de sites offrent des présentations bilingues ou trilingues, voire multilingues, pour toucher un public le plus large possible. Le site du quotidien belge *Le Soir*, par exemple, offre une présentation du journal en six langues: français, allemand, anglais, espagnol, italien et néerlandais. Le Club des poètes, lui, présente son site en anglais, en espagnol et en portugais.

Robert Ware est le créateur de OneLook Dictionaries, qui permet une recherche rapide dans 2.058.544 mots appartenant à 425 dictionnaires couvrant plusieurs domaines: affaires, informatique et Internet, médecine, religion, sciences et techniques, sports, généralités et argot. Dans son courrier électronique du 2 septembre 1998, il écrivait:

"Un fait intéressant s'est produit dans 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 courriers électroniques. Soudain, cela m'a frappé... Le logiciel avait été écrit à 40 km de là, mais c'était une personne située de l'autre côté de la planète qui m'aidait. Les distances et les considérations géographiques 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. Internet a supprimé une barrière, celle de la distance, mais il subsiste la barrière de la langue, bien réelle.

Il semble qu'Internet propulse simultanément les gens dans deux directions différentes. Internet, anglophone à l'origine, relie les gens dans le monde entier. Par là même il favorise une langue commune pour communiquer. Mais il crée aussi des contacts entre des personnes de langue différente et permet ainsi le développement d'un intérêt plus grand pour le multilinguisme. Si une langue commune est appréciable, elle ne remplace en aucun cas cette nécessité.

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 Internet fournit la possibilité de les rendre disponibles rapidement et à bon marché."

## 2.2. Le français face au géant anglais

Comme on l'a vu plus haut, le fort pourcentage de l'anglais - 100% à l'origine, environ 80% maintenant - est à la baisse. Comme il s'étend progressivement à tous les continents et à de nombreux pays non anglophones, Internet devient peu à peu multilingue. Le français est pris ici comme "exemple de langue non anglophone".

Les sites francophones ont enregistré une forte expansion depuis 1996. Au début de 1998, les Québécois attendaient de pied ferme l'arrivée en masse de sites français, surtout dans le domaine du commerce électronique. Le 10 février 1998, lors d'un entretien avec le cyberquotidien Multimédium, Louise Beaudouin, ministre de la Culture et des Communications du Québec, déclarait: "J'attendais depuis deux ans que la France se réveille. Aujourd'hui, je ne m'en plaindrai pas. A cette date, le Québec (6 millions d'habitants) proposait plus de sites web que la France (60 millions d'habitants). La ministre attribuait le retard de la France à deux facteurs: d'une part les tarifs élevés du téléphone et du Minitel, d'autre part les transactions commerciales possibles sur le Minitel depuis plusieurs années, ce qui a ralenti l'expansion du commerce électronique sur Internet.

"En voulant trop en faire une affaire nationale, qui exprimerait aussi par ailleurs l'antipathie qu'ils ont envers les Anglais, les Français ont tendance à freiner la propagation de leur culture. Cela est très regrettable, lit-on le 7 novembre 1996 dans *Yomiyuri Shimbun*, le plus grand quotidien japonais avec ses dix millions d'exemplaires (cité dans un article de Pierre Perroud, créateur de la cyberbibliothèque Athena). Ce cliché a-t-il jamais été vrai, si ce n'est pendant la Guerre de cent ans?

Plus optimiste, Tim Berners-Lee, créateur du World Wide Web, déclarait cependant en décembre 1997 à Pierre Ruetschi dans la Tribune de Genève: "Pourquoi les francophones ne mettent-ils pas davantage d'informations sur le Web? Est-ce qu'ils pensent que personne ne veut la lire, que la culture française n'a rien à offrir? C'est de la folie, l'offre est évidemment énorme." Ces remarques sont-elles une critique, un encouragement, ou les deux? On aurait pu les comprendre il y a deux ou trois ans, mais il n'est pas sûr qu'elles soient encore de mise aujourd'hui, il suffit de naviguer sur le Web francophone pour s'en rendre compte.

Le français n'est pas seulement la langue de la France, du Québec et d'une partie de la Belgique et de la Suisse. Il est la langue officielle de 49 états, dont un certain nombre de pays africains, ce qui représente 500 millions de personnes. Un exemple parmi d'autres de la coopération francophone est l'Agence de la francophonie. Créée en 1970 pour regrouper 21 états francophones, elle en comptait 47 en 1997. "Instrument de coopération multilatérale née d'un idéal, celui de créer une communauté qui fasse entendre sa voix dans le concert des nations, [elle] participe aujourd'hui à l'avènement d'un Secrétariat général de la Francophonie."

S'il est la langue des pays francophones, le français est aussi la deuxième langue utilisée dans les organisations internationales. Malgré la pression anglophone, réelle ou supposée selon les cas, des francophones veillent à ce que leur langue ait sa place en Europe et dans le monde, au même titre que les autres grandes langues de communication que sont l'anglais, l'arabe, le chinois et l'espagnol. Là aussi, l'optique est aussi bien la défense d'une langue que le respect du multilinguisme et de la diversité des peuples.

C'est l'Annuaire de l'UREC (UREC: Unité réseaux du CNRS - Centre national de la recherche scientifique) qui, en France, a été le premier annuaire de sites web en langue française. Travail de pionnier, il a permis aux cybernautes francophones d'une part de se familiariser avec le Web sans se noyer dans la masse d'informations mondiale, d'autre part de connaître les sites qui petit à petit fleurissaient dans le monde francophone. Créé au début de 1994, il a d'abord recensé les sites académiques, puis son contenu est devenu plus généraliste.

Comme l'expliquait Claude Gross sur le site à l'automne 1997, la gestion de l'annuaire est ensuite devenue très difficile du fait de l'accroissement constant du nombre de sites web, et notamment de sites commerciaux. Par la suite, d'autres annuaires ont vu le jour, dont certains débutés avec l'aide de l'UREC. En juillet 1997, considérant que la mission qu'elle s'était donnée était accomplie, l'UREC a donc arrêté la mise à jour de cet annuaire généraliste. Il est maintenant remplacé par un annuaire spécialisé consacré à l'enseignement supérieur et à la recherche.

Nombreux sont ceux qui travaillent à une meilleure représentation du français sur le Web, en prônant un Web francophone intégré dans un Web international et multilingue tenant compte de la diversité des langues en Europe et dans le monde.

La Délégation générale à la langue française (DGLF) s'est donnée plusieurs missions: veiller à la promotion et à l'emploi du français en France, favoriser son utilisation comme langue de communication internationale, et développer le plurilinguisme, garant de la diversité culturelle. La rubrique: "France langue" propose trois listes de diffusion consacrées à la langue française: "France langue", "France langue assistance" et "France langue technologies". Gérée et modérée par la DGLF, France langue se veut "un lieu convivial d'échanges d'informations et d'idées (manifestations, colloques, publications, etc.), ainsi qu'un lieu de discussion sur les thèmes liés à la langue française, à la diversité linguistique, à la dynamique des langues, à la politique linguistique", et elle accueille toutes les questions d'ordre linguistique (grammaire, orthographe, usage, etc.).

La DGLF a également mené plusieurs actions pour assurer la place du français sur les nouveaux réseaux, notamment l'édition de guides techniques sur l'utilisation dans les logiciels des caractères typographiques et des accents propres à la langue française, ce en liaison avec l'AFNOR (Association française de normalisation), ou la francisation de ceux qui sont commercialisés en France.

Le site de la Maison de la Francité, association belge sans but lucratif subventionnée depuis 1976 par la Commission communautaire française, souhaite présenter la réalité socio-linguistique de Bruxelles, seconde capitale internationale de langue française après Paris, tout en agissant pour la défense et la promotion de la langue française à Bruxelles et au sein de la Communauté française Wallonie-Bruxelles.

Au Québec, l'Office de la langue française (OLF), organisme gouvernemental chargé d'assurer la promotion du français, veille à l'implantation et au maintien du français dans les milieux de travail et des affaires et dans l'administration. Il définit et conduit la politique québécoise en matière de linguistique et de terminologie.

Les instances politiques ont également contribué à favoriser l'accès des autoroutes de l'information aux francophones. En application de la Résolution sur la société de l'information adoptée par les chefs d'Etat et de gouvernement à Cotonou (Bénin) en décembre 1995, la Conférence des ministres francophones chargés des inforoutes s'est déroulée à Montréal (Québec) du 19 au 21 mai 1997. Datée du 21 mai 1997, la Déclaration de Montréal proposait de "développer une aire francophone d'éducation, de formation et de recherche; soutenir la création et la circulation de contenus francophones et contribuer à la sauvegarde et à la valorisation des patrimoines; encourager la promotion de l'aire francophone de développement économique; mettre en place une vigie francophone

(veille active); sensibiliser prioritairement la jeunesse ainsi que les utilisateurs, les producteurs et les décideurs; assurer la présence et la concertation des francophones dans les instances spécialisées."

La propagation d'une langue passe aussi par l'étude dynamique de celle-ci. Internet ouvre des horizons sans précédent sur l'utilisation de bases de données linguistiques et les possibilités de recherche textuelle, témoin le site de l'Institut national de la langue française (INaLF), qui présente ses recherches sur la langue française, notamment le discours littéraire des 14<sup>e</sup> au 20<sup>e</sup> siècles (contenu, sémantique, thématique), la langue courante (langue écrite, langue parlée, argot), le discours scientifique et technique et ses ressources terminologiques.

Christiane Jadelot, ingénieur d'études à l'INaLF-Nancy, expliquait dans son courrier électronique du 8 juin 1998:

"Les premières pages sur l'INaLF ont été mises sur l'Internet au milieu de l'année 96, à la demande de Robert Martin, directeur de l'INaLF. Je peux en parler, car j'ai participé à la mise sous Internet de ces pages, avec des outils qui ne sont pas comparables à ceux que l'on utilise aujourd'hui. J'ai en effet travaillé avec des outils sous UNIX, qui n'étaient pas très faciles d'utilisation. Nous avons peu d'expérience de la chose, à l'époque, et les pages étaient très verbeuses. Mais la direction a senti la nécessité urgente de nous faire connaître par l'Internet, que beaucoup d'autres entreprises utilisaient déjà pour promouvoir leurs produits. Nous sommes en effet "Unité de recherche et de service" et nous avons donc à trouver des clients pour nos produits informatisés, le plus connu d'entre eux [étant] la base textuelle FRANTEXT. Il me semble que la base FRANTEXT était déjà sur Internet [depuis début 1995], ainsi qu'une maquette du tome 14 du TLF [Trésor de la langue française]. Il était donc nécessaire de faire connaître l'ensemble de l'INaLF par ce moyen. Cela correspondait à une demande générale."

Organisme à but non lucratif, le Centre d'expertise et de veille Inforoutes et Langues (CEVEIL) existe depuis 1995. Sa création découle du désir du gouvernement du Québec de mieux cerner la problématique de l'utilisation et du traitement des langues sur les inforoutes - dans une optique plus spécifiquement francophone - par le biais d'activités de veille stratégique et la création d'un réseau d'échanges et d'expertise.

Le CEVEIL s'est donné les objectifs suivants : sensibiliser les entreprises québécoises aux bénéfices associés à une gestion électronique et stratégique de leurs savoirs; transférer aux entreprises québécoises les pratiques utiles pour assurer cette gestion électronique et stratégique; mettre en évidence le plurilinguisme et l'importance des normes et standards, dans un contexte d'enjeux technologiques et stratégiques reliés à la gestion des savoirs; valoriser l'expertise québécoise et susciter des collaborations entre les différents acteurs, au Québec ou ailleurs.

Guy Bertrand, directeur scientifique du CEVEIL, et Cynthia Delisle, consultante, ont répondu à mes questions dans leur courrier électronique du 23 août 1998.

ML: "Comment voyez-vous le multilinguisme sur le Web?"

CEVEIL: "Le multilinguisme sur Internet est la conséquence logique et naturelle de la diversité des populations humaines. Dans la mesure où le Web a d'abord été développé et utilisé aux Etats-Unis, il n'est guère étonnant que ce médium ait commencé par être essentiellement anglophone (et le demeure actuellement). Toutefois, cette situation commence à se modifier et le mouvement ira en s'amplifiant, à la fois parce que la plupart des nouveaux usagers du Réseau n'auront pas l'anglais comme langue maternelle et parce que les communautés déjà présentes sur le Web accepteront de moins en moins la "dictature" de la langue anglaise et voudront exploiter Internet dans leur propre langue, au moins partiellement.

On peut prévoir que l'on arrivera sans doute, d'ici quelques années, à une situation semblable à ce qui prévaut dans le monde de l'édition en ce qui a trait à la répartition des différentes langues. Ceci signifie néanmoins que seul un nombre relativement restreint de langues seront représentées (comparativement aux quelques milliers d'idiomes qui existent). Dans cette optique, nous croyons que le Web devrait chercher, entre autres, à favoriser un renforcement des cultures et des langues minoritaires, en particulier pour les communautés dispersées.

Enfin, l'arrivée de langues autres que l'anglais sur Internet, si elle constitue un juste rééquilibrage et un enrichissement indéniable, exaspère évidemment le besoin d'outils de traitement linguistique aptes à gérer efficacement cette situation, d'où la nécessité de poursuivre les travaux de recherche et les activités de veille dans des secteurs comme la traduction automatique, la normalisation, le repérage de l'information, la condensation automatique (résumés), etc."

ML: "Quel a été le bénéfice de l'utilisation d'Internet pour votre organisme?"

CEVEIL: "Mentionnons, tout d'abord, que l'existence du Web constitue en soi une des raisons d'être du CEVEIL, puisque nous concentrons nos activités principalement autour de la thématique de l'utilisation et du traitement des langues sur Internet.

Par ailleurs, le Web est notre principal terrain de cueillette d'information sur les thématiques qui nous préoccupent. Nous procédons notamment à une fréquentation assidue des sites abordant nos thématiques de travail - plus particulièrement des sites diffusant des nouvelles quotidiennes et/ou hebdomadaires. A ce niveau, on peut affirmer sans hésitation que nous exploitons davantage Internet que les diverses ressources écrites disponibles pour réaliser nos activités.

Dans un ordre d'idées un peu différent, nous utilisons abondamment le courriel pour entretenir des relations avec les intervenants du milieu et ainsi obtenir des informations et mener à bien divers projets. Le CEVEIL est un "structure-réseau" qui survivrait difficilement sans Internet pour relier toutes les personnes impliquées.

Enfin, il convient de signaler que le Web constitue également notre plus important outil pour la diffusion de nos produits aux clientèles-cibles: envoi de bulletins électroniques de nouvelles à nos abonnés, création d'un périodique électronique, diffusion d'information et de documents via notre site web, etc."

ML: "Comment voyez-vous l'évolution vers un Internet multilingue?"

CEVEIL: "Internet est là pour demeurer; l'apparition de langues autres que l'anglais sur ce médium constitue également un phénomène irréversible. Il sera donc nécessaire de tenir compte de ces nouvelles réalités aux points de vue économique, social, politique, culturel, etc. Des secteurs comme la publicité, la formation professionnelle, le travail en groupes et en réseaux, la gestion des connaissances devront évoluer en conséquence. Cela nous ramène, tel que nous l'avons mentionné plus haut, à la nécessité de développer des technologies et des outils vraiment performants qui faciliteront les échanges dans un Village global terriblement plurilingue..."

### 2.3. Diversité des langues dans l'Union européenne

Henri Slettenhaar, professeur à la Webster University (Genève, Suisse), est un Européen trilingue. Il est hollandais, il enseigne l'informatique en anglais et il parle couramment le français puisqu'il vit en France. Dans son courrier électronique du 21 décembre 1998, il expliquait:

"Je vois le multilinguisme comme un facteur fondamental. Les communautés locales présentes sur le Web devraient en tout premier lieu utiliser leur langue pour diffuser des informations. Si elles veulent également présenter ces informations à la communauté mondiale, celles-ci doivent aussi être disponibles en anglais. Je pense qu'il existe un réel besoin de sites bilingues.

[...] 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."

Selon Global Reach, 15% seulement des Européens (qui représentent 500 millions d'habitants) sont de langue maternelle anglaise, et 28% seulement parlent anglais. Le site mentionne aussi une étude récente démontrant que 32% seulement des cybernautes européens consultent le Web anglophone.

Créateur de Euro-Marketing Associates (qui englobe Global Reach), Bill Dunlap, qui défend le commerce électronique européen auprès de ses compatriotes américains, écrivait dans son courrier électronique du 12 décembre 1998, que contrairement à l'Amérique du Nord, "ici en Europe (j'écris de France), les pays sont petits, si bien que, depuis des siècles, une perspective internationale est nécessaire."

De nombreux organismes traitent de multilinguisme, par exemple l'Association européenne pour les ressources linguistiques (ELRA), le European Network in Language and Speech (ELNET) et le programme Multilingual Information Society (MLIS) de la Communauté européenne.

Etablie à Luxembourg en Février 1995, l'Association européenne pour les ressources linguistiques (ELRA) est une association à but non lucratif qui a pour but de fournir une organisation centralisée pour la validation, la gestion et la distribution des ressources et outils en parole, texte et terminologie, et de promouvoir leur utilisation au sein de la communauté européenne de recherche et développement en télématique.

Le European Network in Language and Speech (ELNET) (Réseau européen pour le langage et la parole) regroupe plus d'une centaine d'institutions universitaires et industrielles. L'objectif

technologique commun aux participants de ELSNET est de construire des systèmes multilingues pour la parole et le langage naturel, avec couverture illimitée de la langue parlée et écrite.

Dans son courrier électronique du 23 septembre 1998, Steven Krauwer, coordinateur d'ELSNET, expliquait:

"En tant que citoyen européen je pense que le multilinguisme sur le Web est absolument essentiel. A 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.

Internet est l'instrument que j'utilise le plus pour communiquer avec les autres, et c'est ma source principale d'information. [...] Je compte passer le reste de ma vie professionnelle à utiliser la technologie de l'information pour supprimer ou réduire les barrières des langues."

Lancé en novembre 1996, le programme Multilingual Information Society (MLIS) (Société d'information multilingue) de la Commission européenne soutient la réalisation d'une infrastructure favorisant des ressources linguistiques européennes, mobilise et accroît le développement des industries du langage, et favorise l'utilisation d'outils linguistiques dans le secteur public européen. Les entreprises, les organismes du secteur public, les industries du langage et les citoyens peuvent bénéficier de ce programme.

"Afin de promouvoir 'concrètement' une Europe multilingue, dans un contexte d'évolution rapide et permanent des nouvelles technologies de communication, et d'attirer largement l'intérêt sur ce programme, ce site est accessible dans les 11 langues officielles de l'Union européenne [allemand, anglais, danois, espagnol, finlandais, français, grec, hollandais, italien, portugais et suédois]. Nous renvoyons à des références et à des liens encore plus larges... Nous essayons de fournir autant d'information que possible dans ces onze langues, et nous vous demandons de rester compréhensif pour le retard existant parfois entre la publication d'un document officiel dans sa langue initiale et sa traduction dans les autres langues de l'UE."

Le multilinguisme est l'affaire de tous, témoin cet Appel du Comité européen pour le respect des cultures et des langues en Europe (CERCLE) qui, diffusé dans les onze langues officielles de l'Union européenne, défend "une Europe humaniste, plurilingue et riche de sa diversité culturelle". Il propose aux réviseurs du Traité de l'Union européenne douze amendements prenant en compte le respect des cultures et des langues:

"La diversité et le pluralisme linguistiques ne sont pas un obstacle à la circulation des hommes, des idées et des marchandises ou services, comme veulent le faire croire certains, alliés objectifs, conscients ou non, de la culture et de la langue dominantes. C'est l'uniformisation et l'hégémonie qui sont un obstacle au libre épanouissement des individus, des sociétés et de l'économie de l'immatériel, source principale des emplois de demain. Le respect des langues, à l'inverse, est la dernière chance pour l'Europe de se rapprocher des citoyens, objectif toujours affiché, presque jamais mis en pratique. L'Union doit donc renoncer à privilégier la langue d'un seul groupe."

Dans Language Futures Europe, Paul Treanor propose des liens sur la politique linguistique, le multilinguisme, les structures linguistiques globales, et la domination de l'anglais. Le site débute par un commentaire sur les structures linguistiques. Il offre des textes et des essais, des sections consacrées à la politique européenne, aux politiques nationales et aux sites de recherche, et des documents sur le "mouvement monolingue" émergent aux Etats-Unis.

Dans son courrier électronique du 18 août 1998, Paul Treanor exposait ses réflexions sur les questions posées:

"Vous parlez du Web au singulier. Comme vous l'avez sans doute lu, je pense que "Le Web" est un concept politique, non technologique. Une civilisation est possible avec des ordinateurs très sophistiqués mais pas d'interconnexion. L'idée qu'il devrait exister "Un Web" est dérivée de la tradition libérale du marché unique ouvert, de préférence mondial.

J'ai déjà suggéré qu'Internet devrait être tout simplement découpé, et que l'Europe devrait couper ses liens avec les Etats-Unis, et construire un autre Net spécifique et incompatible avec le premier. Dès qu'on envisage la possibilité de Nets multiples, les thèmes correspondant aux différents chapitres de votre étude sont souvent hors de propos. Rappelez-vous qu'il y a quinze ans tout le monde pensait qu'il n'y aurait qu'un émetteur de télévision à l'échelle mondiale, CNN. Or il existe maintenant des chaînes de télévision nationales françaises, allemandes ou espagnoles. La réponse à votre question est donc que



l'entité "Un Web" sera de toute manière divisée, probablement en quatre parties:

- a) un Net interne aux Etats-Unis et au Canada, avec nombre des caractéristiques actuelles;
- b) des Nets nationaux séparés, avec des liens limités avec l'extérieur;
- c) un nouveau Net général pour relier entre eux les Nets de la catégorie 2;
- d) et peut-être un Net spécifique à l'Union européenne.

Comme vous voyez, cette structure est parallèle à celle qui existe en géopolitique. Toute l'infrastructure des télécommunications a suivi des modèles similaires.

Je pense qu'il n'est pas possible d'approcher le Web de la manière neutre et apolitique envisagée dans votre étude. La politique actuelle de l'Union européenne prétend être neutre, mais elle supporte en fait le développement de l'anglais comme langue de contact pour communiquer."

### **3. RESSOURCES LINGUISTIQUES**

[Dans ce chapitre:]

[3.1. Sites indexant des ressources linguistiques / 3.2. Répertoires linguistiques / 3.3. Dictionnaires et glossaires / 3.4. Bases de données textuelles / 3.5. Bases de données terminologiques]

#### **3.1. Sites indexant des ressources linguistiques**

Préparé par le programme Télématique pour les bibliothèques de la Commission européenne, Multilingual Tools and Services (Outils et services multilingues) propose une série de dictionnaires, instruments d'aide multilingues, projets, moteurs de recherche par langue, banques de données terminologiques, thésaurus et systèmes de traduction.

Créée par Tyler Chambers en mai 1994, The Human-Languages Page (La page des langues humaines) est un catalogue détaillé de 1.800 ressources linguistiques dans plus de 100 langues différentes. Les différentes sections sont: langues et littérature, écoles et institutions, ressources linguistiques, produits et services, organismes, emplois et stages, dictionnaires et cours de langues.

Tyler Chambers mène aussi un autre projet relatif aux langues, l'Internet Dictionary Project (Projet de dictionnaire Internet). Sur le site web, il explique:

"Le but de l'Internet Dictionary Project est de créer des dictionnaires de traduction grâce à l'aide des internautes. Ce site permet aux individus du monde entier de 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. [...]

The Internet Dictionary Project a débuté en 1995 pour combler une lacune et procurer à la communauté d'Internet et à tous ceux qui s'intéressent à l'informatique des dictionnaires de traduction gratuits. Non seulement il est très utile d'avoir immédiatement accès à des dictionnaires via 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 Internet et susciter d'autres projets qui seront plus bénéfiques que des revenus purement financiers."

Tyler Chambers a répondu à mes questions dans son courrier électronique du 14 septembre 1998.

ML: "Comment voyez-vous le multilinguisme sur le Web?"

TC: "Le multilinguisme sur le Web était inévitable bien avant que ce médium ne se développe vraiment. Mon premier vrai contact avec Internet 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. Suite à l'expérience acquise maintenant par les navigateurs et les cybernautes, je ne pense pas qu'il existe une langue vivante qui

ne soit pas 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. A ma connaissance, très peu de jeux de caractères ne sont pas disponibles en ligne: les navigateurs ont maintenant la possibilité de visualiser les caractères romains, asiatiques, cyrilliques, grecs, turcs, etc. Accent Software a un produit appelé "Internet avec accents" qui prétend être capable de visualiser plus de 30 codages différents. S'il existe encore des obstacles à la diffusion d'une langue spécifique sur le Web, ceci ne devrait pas durer."

ML: "Quel a été le bénéfice de l'utilisation d'Internet dans votre vie professionnelle?"

TC: "Ma vie professionnelle est en ce moment complètement distincte de mon activité sur Internet. Je suis un programmeur/technicien informatique, je trouve cela stimulant et cela me permet de payer les factures. Mon activité en ligne a été de rendre l'information linguistique accessible à davantage de gens par le biais de deux de mes projets sur le Web. 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. Internet m'a permis de toucher des millions de personnes et de les aider à trouver ce qu'elles cherchaient, chose que je suis heureux de faire. Je suis devenu aussi une sorte de célébrité, ou au moins quelqu'un de familier dans certains cercles. Je viens de découvrir qu'un de mes projets est brièvement mentionné dans les éditions asiatique et internationale de Time Magazine. Dans l'ensemble, je pense que le Web a été 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? Les médias de communication rendent le monde plus petit en rapprochant les gens; je pense que le Web est le premier médium - bien plus que le courrier, le télégraphe, le téléphone, la radio ou la télévision - à réellement permettre à l'utilisateur moyen de franchir les frontières nationales et culturelles. Israël n'est plus à des milliers de kilomètres, mais seulement à quelques clics de souris. Notre monde est maintenant suffisamment petit pour tenir sur un écran d'ordinateur."

ML: "Comment voyez-vous l'évolution vers un Internet multilingue?"

TC: "Comme je l'ai dit plus haut, je pense que l'avenir d'Internet réside dans davantage de multilinguisme et d'exploration et de compréhension multiculturelles que nous n'en avons jamais vu. Cependant Internet sera seulement le médium au travers duquel l'information circule. Comme le papier qui sert de support au livre, Internet lui-même augmente très peu le contenu de l'information. Par contre il augmente énormément la valeur de celle-ci dans la capacité qu'il a à communiquer cette information. Dire qu'Internet aiguillonne le multilinguisme est à mon sens une opinion fautive. C'est la communication qui aiguillonne le multilinguisme et l'échange multiculturel, Internet est seulement le mode de communication le plus récent rendu accessible aux gens plus ou moins ordinaires. Internet a un long chemin à parcourir avant d'être omniprésent dans le monde entier, mais il est vraisemblable que lui-même ou un médium de la même lignée y arrive. Les langues deviendront encore plus importantes qu'elles ne le sont quand 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), mais je ne sais pas si ceci mènera à un renforcement des attaches linguistiques ou à une fusion des langues jusqu'à ce qu'il n'en subsiste plus que quelques-unes ou même une seule. Une chose qui m'apparaît certaine est qu'Internet sera toujours la marque de notre diversité, y compris la diversité des langues, même si cette diversité diminue. Et c'est une des choses que j'aime au sujet d'Internet, c'est un exemple à l'échelle mondiale du dicton: 'Cela n'a pas vraiment disparu tant que quelqu'un s'en souvient.' Et les gens se souviennent."

Depuis ses débuts en 1989, le CTI (Computer in Teaching Initiative) Centre for Modern Languages (Centre pour l'utilisation des ordinateurs dans l'enseignement des langues modernes) est inclus dans l'Institut des langues de l'Université d'Hull (Royaume-Uni) et vise à promouvoir l'utilisation des ordinateurs dans l'apprentissage et l'enseignement des langues. Le Centre procure des informations sur la manière dont l'apprentissage des langues assisté par ordinateur peut être effectivement intégré à des cours existants en offrant un soutien aux professeurs qui utilisent, ou souhaitent utiliser, l'informatique dans l'enseignement qu'ils dispensent.

June Thompson, responsable du Centre, a répondu à mes questions dans son courrier électronique du 14 décembre 1998.

ML: "Comment voyez-vous le multilinguisme sur Internet?"

JT: "Avec Internet, on a la possibilité d'accroître l'utilisation des langues étrangères, et notre organisation ne soutient absolument pas la suprématie de l'anglais en tant que langue d'Internet. Une intéressante conférence sur ce sujet a été donnée par Madanmohan Rao à la Conférence WorldCALL de Melbourne en juillet 1998. [Voir la présentation de l'ouvrage à paraître suite à la conférence.]

ML: "Quel a été le bénéfice de l'utilisation d'Internet dans votre activité?"

JT: "L'utilisation d'Internet a apporté une nouvelle dimension à notre tâche qui consiste à soutenir les professeurs de langue dans l'utilisation de la technologie correspondante."

ML: "Comment voyez-vous l'évolution vers un Internet multilingue?"

JT: "Je pense que dans un avenir proche l'utilisation de supports linguistiques sur Internet va continuer à se développer en même temps que d'autres activités liées à la technologie (par exemple l'utilisation de CD-ROM - certains établissements n'ont pas suffisamment de matériel informatique en réseau). A l'avenir il me semble que l'utilisation d'Internet jouera un rôle plus grand, mais seulement si ces activités sont à caractère pédagogique. Notre organisme travaille étroitement avec le WELL project [Web Enhanced Language Learning - Apprentissage des langues grâce au Web], qui se consacre à ces problèmes."

Hébergé par le CTI Centre for Modern Languages et l'Université d'Hull (Royaume-Uni), EUROCALL (European Association for Computer Assisted Language Learning) (Association européenne pour l'apprentissage des langues assisté par ordinateur) regroupe des professionnels de l'enseignement des langues exerçant en Europe et dans le monde entier. Ses objectifs sont de favoriser l'utilisation des langues étrangères en Europe, ainsi qu'une vision européenne de l'utilisation de la technologie pour l'apprentissage des langues et l'élaboration et la diffusion d'un matériel de qualité.

EUROCALL soutient aussi deux groupes: CAPITAL (Computer Assisted Pronunciation Investigation Teaching and Learning) (Investigation, enseignement et apprentissage de la prononciation assistés par ordinateur), qui est un groupe de chercheurs et de praticiens souhaitant utiliser l'informatique dans le domaine de la prononciation pris dans le sens le plus large du terme, et WELL (Web Enhanced Language Learning) (Apprentissage des langues grâce au Web), qui procurera l'accès à des ressources web de qualité dans douze langues différentes. Sélectionnées et décrites par des experts, ces ressources seront complétées par des informations et des exemples sur la manière de les utiliser pour l'enseignement et l'apprentissage d'une langue.

Internet Resources for Language Teachers and Learners (Ressources Internet pour l'enseignement et l'apprentissage des langues) propose plusieurs catégories de liens: ressources linguistiques générales (centres et départements, dictionnaires et grammaires, listes de discussion, apprentissage des langues à distance, caractères, compte-rendus, linguistique, listes et index, divers, journaux et périodiques, organismes, ressources, sites linguistiques multilingues, moteurs de recherche et index, sites linguistiques commerciaux (audiovisuels, écoles de langue, ressources et répertoires, logiciels).

Géré par l'Institut des sciences phonétiques d'Amsterdam (Pays-Bas), Speech on the Web (Parole sur le Web) est un répertoire de liens organisés en différentes sections: congrès, réunions et ateliers, liens et listes, phonétique et parole, traitement de la langue naturelle, sciences cognitives et intelligence artificielle, linguistique computationnelle, dictionnaires, lettres d'information électroniques, revues et publications.

Travlang est un site dédié à la fois aux voyages et aux langues. Créé par Michael C. Martin en 1994 sur le site de son université alors qu'il était étudiant en physique (il est maintenant chercheur au Lawrence Berkeley National Laboratory, en Californie), Foreign Languages for Travelers (Langues étrangères pour les voyageurs), inclus dans Travlang en 1995, donne la possibilité d'apprendre 60 langues différentes sur le Web. Translating Dictionaries (Dictionnaires de traduction) donne accès à des dictionnaires gratuits dans diverses langues (afrikaans, allemand, danois, espagnol, espéranto, finnois, français, frison, hollandais, hongrois, italien, latin, norvégien, portugais et tchèque). Le site offre aussi de nombreux liens vers des dictionnaires de langue, services de traduction, écoles de langue, librairies multilingues, etc.

Michael C. Martin a répondu à mes questions dans son courrier électronique du 25 août 1998.

ML: "Comment voyez-vous le multilinguisme sur le Web?"

MCM: "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 souhaitent se sentir en contact avec d'autres parties du monde."

ML: "Quel a été le bénéfice de l'utilisation d'Internet dans votre activité?"

MCM: "Et bien, nous en avons fait une petite affaire! Internet est vraiment un outil important pour communiquer avec des gens avec lesquels vous n'auriez pas l'occasion de dialoguer autrement. J'apprécie vraiment la collaboration générale qui a rendu possibles les pages de Foreign Languages for Travelers (Langues étrangères pour les voyageurs)."

ML: "Comment voyez-vous l'évolution vers un Internet multilingue?"

MCM: "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 Internet au monde non anglophone."

La LINGUIST List est la section linguistique de la WWW Virtual Library. Elle donne une série complète de liens sur les ressources linguistiques: profession (conférences, associations linguistiques, programmes, etc.), recherche et soutien à la recherche (articles, résumés de mémoires, projets, bibliographies, dossiers, textes), publications, pédagogie, ressources linguistiques (langues, familles linguistiques, dictionnaires, information régionale), et soutien informatique (polices de caractères et logiciels).

Helen Dry, modératrice de la LINGUIST List, expliquait dans son courrier électronique du 18 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: 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 qui présente 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."

Géré par le Yamada Language Center de l'Université d'Oregon (USA), Yamada WWW Language Guides (Guides Yamada de langues sur le WWW) est un répertoire de ressources linguistiques par famille géographique et par ordre alphabétique. Il comprend des organismes, des écoles de langues, du matériel d'enseignement, des références culturelles, et des liens WWW.

Language today (La langue aujourd'hui) est un nouveau magazine pour ceux qui travaillent dans les langues appliquées: traducteurs, interprètes, terminologues, lexicographes et rédacteurs techniques. C'est une réalisation commune de Logos, qui procure le site web, et Praetorius, service d'expertise britannique qui suit l'actualité des développements concernant les langues appliquées. Le site procure des liens vers des associations de traducteurs, des écoles de langues et des dictionnaires.

Geoffrey Kingscott, directeur général de Praetorius, a répondu à mes questions dans son courrier électronique du 4 septembre 1998.

ML: "Comment voyez-vous le multilinguisme sur le Web?"

GK: "Les caractéristiques propres au Web sont la multiplicité de générateurs de sites et le bas prix de l'émission de messages. Ceci favorisera donc le multilinguisme au fur et à mesure du développement du Web. Comme celui-ci a vu le jour aux Etats-Unis, il est encore principalement en anglais, mais ce n'est qu'un phénomène temporaire. Pour expliquer ceci plus en détail, je dirais que quand nous comptions sur l'imprimé ou l'audiovisuel (film, télévision, radio, vidéo, cassettes), l'information ou le divertissement que nous attendions dépendait d'agents (éditeurs, stations de télévision ou de radio, producteurs de cassettes ou de vidéos) qui devaient subsister commercialement et, dans le cas de la radiotélédiffusion du service public, avec de sévères contraintes budgétaires. Ceci signifie que la quantité de clients est primordiale, et détermine la nécessité de langues autres que l'omniprésent anglais. Ces contraintes disparaissent avec le Web. Pour ne donner qu'un exemple mineur tiré de notre expérience, nous publions la version imprimée de Language Today uniquement en anglais, dénominateur commun de nos lecteurs. Quand nous utilisons un article qui était originellement dans une langue autre que l'anglais, ou que nous relatons un entretien mené dans une langue autre que l'anglais, nous le traduisons en anglais et nous ne publions que la version anglaise, pour la raison suivante: le nombre de pages que nous pouvons imprimer est limité, et déterminé en fonction de notre clientèle (annonceurs et abonnés). Par contre, dans notre version Web, nous proposons aussi la version originale."

ML: "Quel a été le bénéfice de l'utilisation d'Internet dans votre activité?"

GK: "Internet n'a pas apporté de changement majeur dans notre société. C'est un médium de plus plutôt qu'un médium visant à remplacer les autres."

ML: "Quelle est l'évolution prévue dans les prochaines années?"

GK: "Nous continuerons d'avoir un site web pour notre société, et de publier une version de notre revue sur le Web, mais ceci ne sera qu'un secteur de notre travail. Nous utilisons Internet comme une source d'information que nous distillons ensuite à nos lecteurs, qui autrement seraient confrontés au problème majeur du Web: faire face à un flux incontrôlé d'informations."

### 3.2. Répertoires linguistiques

The Ethnologue est la version électronique de The Ethnologue, 13e éd., (éditrice: Barbara F. Grimes, avec la collaboration de Richard S. Pittman and Joseph E. Grimes), publié en 1996 par le Summer Institute of Linguistics (Dallas, Texas, USA). Ce catalogue de plus de 6.700 langues parlées dans 228 pays est accessible au moyen de deux outils de recherche: le Ethnologue Name Index (Index des noms de l'Ethnologue), qui donne la liste des noms de langues et de dialectes et de leurs synonymes, et le Ethnologue Language Family Index (Index des familles linguistiques de l'Ethnologue), qui organise les langues en fonction des familles linguistiques.

Barbara F. Grimes, éditrice de The Ethnologue, écrivait dans son courrier électronique du 18 août 1998:

"Les pages web multilingues sont de plus en plus utiles, mais elles sont plus onéreuses à gérer. Nous avons eu des demandes nous demandant l'accès à The Ethnologue dans plusieurs autres langues, mais nous n'avons pas le personnel ni les fonds pour la traduction ou la réactualisation, indispensable puisque notre site est constamment mis à jour.

Internet nous est utile, c'est un outil pratique qui apporte un complément à notre travail. Nous l'utilisons principalement pour le courrier électronique.

C'est aussi un moyen commode pour mettre notre documentation à la disposition d'une audience plus large que celle de The Ethnologue imprimé.

D'un autre côté, The Ethnologue sur Internet n'atteint en fait qu'une audience limitée disposant d'ordinateurs. Or, dans les personnes que nous souhaitons atteindre, nombreux sont ceux qui n'ont pas accès à des ordinateurs. Je pense particulièrement aux habitants du dit 'Tiers-monde'."

Créé en décembre 1995 par Yoshi Mikami, de Asia Info Network, The Languages of the World by Computers and the Internet (Les langues du monde par ordinateur ou sur Internet), communément appelé Logos Home Page ou Kotoba Home Page, donne, pour chaque langue, un bref historique, les caractéristiques, le système d'écriture, le jeu de caractères et la configuration du clavier pour l'utilisation de l'ordinateur et d'Internet dans la langue donnée.

Dans son courrier électronique du 17 décembre 1998, Yoshi Mikami écrivait:

"Ma langue maternelle est le japonais. Comme j'ai suivi mes études de troisième cycle aux Etats-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. A 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 de toutes ces langues, ainsi que les caractéristiques propres à chaque langue et à sa phonétique. 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 sous le titre: The Multilingual Web Guide (édition japonaise), le premier livre au monde sur un tel sujet.

Il y a des milliers d'années de cela, en Egypte, 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 Etat a adopté plus ou moins une seule langue de communication. A mon avis, 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!"

Disponible sur le site de l'Institut Sabhal Mór Ostaig (Ile de Skye, Ecosse) et géré par Caoimhín P. Ó Donnáile, European Minority Languages est une liste de langues minoritaires - ou rendues minoritaires - disponible par ordre alphabétique et par famille linguistique. Ce site procure aussi des liens avec des sites du même type dans le monde entier.

Caoimhín P. Ó Donnáile écrivait dans son courrier électronique du 18 août 1998:

"Internet a contribué et contribuera au développement fulgurant de l'anglais comme langue mondiale.

Internet peut grandement aider les langues minoritaires. Ceci ne se fera pas tout seul, mais seulement si les gens choisissent de défendre une langue.

Le Web est très utile pour dispenser des cours de langues, et la demande est grande.

La norme UNICODE (ISO 10646) pour les jeux de caractères est très importante et elle va grandement favoriser le multilinguisme sur le Web."

### 3.3. Dictionnaires et glossaires

Les dictionnaires en ligne sont de plus en plus nombreux. Voici trois exemples (francophone, anglophone et multilingue).

Le Dictionnaire francophone en ligne, corpus lexical du Dictionnaire universel francophone, est publié conjointement par Hachette et l'Agence universitaire de la Francophonie (AUPELF-UREF), qui expliquent sur le site:

"Voici enfin présentés, 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. Voici les termes de la flore et de la faune qui apportent vie et couleur à cet ouvrage. Ce dictionnaire s'adresse à tous ceux qui, à travers le monde, veulent voir dans la Francophonie une réalité que l'on désire qualifier de charnelle."

Dans Merriam-Webster Online: the Language Center (Merriam-Webster en ligne: le centre linguistique), un grand éditeur britannique de dictionnaires met en accès libre sa collection de ressources en ligne pour la langue anglaise: définition des mots, orthographe, prononciation, synonymes, exercices de vocabulaire, etc. Les ouvrages de référence disponibles en ligne sont le WWebster Dictionary, le WWebster Thesaurus, le Webster's Third (lexique), le Guide to International Business Communications, le Vocabulary Builder (questions de vocabulaire interactives), et le Barnhart Dictionary Companion (nouveaux mots).

Le Logos Dictionary est un dictionnaire multilingue d'environ 8 millions d'entrées dans toutes les langues. Logos, une société de traduction internationale dont le siège est à Modène (Italie), met en accès libre les outils linguistiques utilisés par ses traducteurs: 200 traducteurs sur place, 2.500 traducteurs en ligne dans le monde, environ 200 textes traités quotidiennement. Outre le Logos Dictionary, ces outils comprennent la Wordtheque, une base de données multilingue de 325 millions de mots provenant de romans, documents techniques et textes traduits et présentés dans leur contexte, Linguistic Resources (Ressources linguistiques), qui regroupe 536 glossaires, et le Universal Conjugator (Conjugaison universelle), qui propose des tableaux de conjugaison dans 17 langues différentes.

Dans Les mots pour le dire, un article du quotidien Le Monde en date du 7 décembre 1997, Annie Kahn écrivait:

"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. A chacun d'enrichir la base, les traducteurs de l'entreprise valident ensuite les traductions proposées."

Logos a été créé par Rodrigo Vergara, un réfugié politique chilien qui a émigré en Italie quand il était étudiant en agronomie pour échapper au régime Pinochet. Aujourd'hui, à 45 ans, il dirige une entreprise de traduction offrant des services dans plus de 35 langues, avec un réseau de 300 traducteurs dans le monde et un chiffre d'affaires de 60 millions de FF.

Dans le même article, Rodrigo Vergara expliquait:

"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."

Les répertoires de dictionnaires sont des outils inappréciables pour les linguistes. Voici par exemple Dictionnaires électroniques, OneLook Dictionaries et A Web of Online Dictionaries.

Préparé par la Section française des Services linguistiques centraux (SLC-f) de la Chancellerie fédérale suisse, Dictionnaires électroniques est classé en cinq secteurs: dictionnaires monolingues, dictionnaires bilingues, dictionnaires multilingues, abréviations et acronymes, et informations géographiques.

Dans son courrier électronique du 14 janvier 1999, Marcel Grangier, responsable de cette section, répondait à mes questions.

ML: "Comment voyez-vous le multilinguisme sur le Web?"

MG: "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. Cette suprématie 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."

ML: "Quel a été le bénéfice de l'utilisation d'Internet dans votre activité?"

MG: "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."

ML: "Comment voyez-vous l'évolution vers un Internet multilingue?"

MG: "La multiplication des langues présentes sur 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."

Disponible depuis avril 1996, OneLook Dictionaries (Dictionnaires tous-dans-un), de Robert Ware, est un moteur de recherche rapide puisant dans plus de 2 millions de mots disponibles dans 425 dictionnaires traitant des domaines suivants: affaires, informatique/Internet, médecine, divers, religion, sciences, sports, technologie, généralités et argot.

Dans son courrier électronique du 2 septembre 1998, Robert Ware expliquait:

"A titre personnel, je suis presque uniquement en contact avec des gens qui ne pratiquent qu'une langue et qui n'ont pas beaucoup de motivation pour développer leurs aptitudes linguistiques. Etre 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."

Le Web of Online Dictionaries (Le Web des dictionnaires en ligne), de Robert Beard, est un index de plus de 800 dictionnaires en ligne dans 150 langues différentes, auquel s'ajoutent d'autres sections: dictionnaires multilingues, dictionnaires anglophones spécialisés, thésaurus et vocabulaires, grammaires en ligne, et outils linguistiques pour non spécialistes.

Robert Beard a répondu à mes questions dans son courrier électronique du 1er septembre 1998.

ML: "Comment voyez-vous le multilinguisme sur le Web?"

RB: "On a d'abord craint que le Web représente un danger pour le multilinguisme, étant donné que le HTML [hypertext markup language] 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, les sites web que je gère montrent 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 en 65 langues différentes. De plus,

comme ceux qui développent les navigateurs manifestent une attention nouvelle pour la diversité des langues dans le monde, ceci favorisera la présence de davantage encore de sites web dans différentes langues."

ML: "Quel a été le bénéfice de l'utilisation d'Internet dans votre vie professionnelle?"

RB: "En tant que professeur de langues, je pense que le Web présente une pléthore de nouvelles ressources disponibles dans la langue étudiée, de nouveaux instruments d'apprentissage (exercices interactifs Java et Shockwave) et de test, qui sont à la disposition des étudiants quand ceux-ci en ont le temps ou l'envie, 24 heures par jour et 7 jours par semaine. Aussi bien pour mes collègues que pour moi, et bien sûr pour notre établissement, Internet nous permet aussi de publier pratiquement sans limitation."

ML: "Comment voyez-vous l'évolution vers un Internet multilingue?"

RB: "Internet nous offrira tout le matériel pédagogique dont nous pouvons rêver, y compris des notes de lecture, exercices, tests, évaluations et exercices interactifs plus efficaces que par le passé parce que reposant davantage sur la notion de communication. 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."

Depuis 1995, à l'initiative du WorldWide Language Institute (Institut des langues du monde entier), NetGlos (The Multilingual Glossary of Internet Terminology) (Le glossaire multilingue de la terminologie d'Internet) est le projet commun d'un certain nombre de traducteurs et autres professionnels. Ce glossaire est préparé dans les langues suivantes: allemand, anglais, chinois, croatien, espagnol, français, grec, hébreu, hollandais/flamand, italien, maori, norvégien et portugais.

Brian King, directeur du WorldWide Language Institute, a répondu à mes questions dans son courrier électronique du 15 septembre 1998.

ML: "Comment voyez-vous le multilinguisme sur le Web?"

BL: "Bien que l'anglais soit la langue la plus importante du Web et d'Internet en général, je pense que le multilinguisme fait inévitablement partie des futures orientations du cyberspace.

Voici quelques-uns des éléments qui, à mon sens, permettront que le Web multilingue devienne une réalité:

a) La popularisation de la technologie de l'information

La technologie des ordinateurs a longtemps été le seul domaine d'une élite 'technicienne', à l'aise à la fois dans des langages de programmation complexes et en anglais, la langue universelle des sciences et techniques. A l'origine, les ordinateurs n'ont jamais été conçus pour manier des systèmes d'écriture ne pouvant être traduits en ASCII. Il n'y avait pas de place pour autre chose que les 26 lettres de l'alphabet anglais dans un système de codage qui, à l'origine, ne pouvait même pas reconnaître les accents aigus et les trémas, sans parler de systèmes non alphabétiques comme le chinois.

Mais la tradition a été bouleversée, et la technologie popularisée. Des interfaces graphiques tels que Windows et Macintosh ont accéléré le processus. La stratégie de marketing de Microsoft a consisté à présenter son système d'exploitation comme facile à utiliser par le client moyen. A l'heure actuelle cette facilité d'utilisation s'est étendue au-delà du PC vers le réseau Internet, si bien que maintenant même ceux qui ne sont pas programmeurs peuvent insérer des applets Java dans leurs pages web sans comprendre une seule ligne de programmation.

b) La compétition des grandes sociétés pour une part de 'marché global'

L'extension de cette popularisation locale est l'exportation de la technologie 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, mais seulement les langues personnelles des 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 high-tech se faisant concurrence pour obtenir les marchés mondiaux a fait de la localisation un secteur en expansion rapide dans le développement des logiciels et du matériel. Le premier pas a été le passage de l'ASCII à l'ASCII



étendu. Ceci signifie que les ordinateurs commençaient à reconnaître les accents et les symboles utilisés dans les variantes de l'alphabet anglais, symboles qui appartenaient le plus souvent aux langues européennes. Cependant une page ne pouvait être affichée qu'en une seule langue à la fois.

#### c) L'innovation technologique

L'innovation la plus récente est UNICODE. Bien qu'il soit encore en train d'évoluer et qu'il ait tout juste été incorporé dans les derniers logiciels, ce nouveau système de codage traduit chaque caractère en 16 octets. Alors que l'ASCII étendu à 8 octets pouvait prendre en compte un maximum de 256 caractères, UNICODE peut prendre en compte plus de 65.000 caractères uniques et il a donc la possibilité de traiter informatiquement tous les systèmes d'écriture du monde.

Les instruments sont maintenant plus ou moins en place. Ils ne sont pas encore parfaits, mais on peut désormais surfer sur le Web en utilisant le chinois, le japonais, le coréen, et de nombreuses autres langues qui n'utilisent pas l'alphabet occidental. Comme Internet s'étend à des parties du monde où l'anglais est très peu utilisé, par exemple la Chine, il est naturel que ce soit le chinois et non l'anglais qui soit utilisé. La majorité des usagers en Chine n'a pas d'autre choix que sa langue maternelle.

Une période intermédiaire précède bien sûr ce changement. Une grande partie de la terminologie technique disponible sur le Web n'est pas encore traduite dans d'autres langues. Et, comme nous nous en sommes rendus compte dans NetGlos, notre glossaire multilingue de la terminologie d'Internet, 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 [fin septembre 1998]. Elle sera sans nul doute un excellent exemple du processus dynamique en cours pour la russification de la terminologie du Web.

#### d) La démocratie linguistique

Dans un rapport de l'UNESCO du début des années 50, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur Internet dans sa langue maternelle pourrait bien être son équivalent à l'Age de l'information. Si Internet doit vraiment devenir le réseau mondial qu'on nous promet, tous les usagers devraient y avoir accès sans problème de langue. Le considérer comme la chasse gardée de ceux qui, par accident historique, nécessité pratique ou privilège politique, connaissent l'anglais, est injuste à l'égard de ceux qui ne connaissent pas cette langue.

#### e) Le commerce électronique

Bien qu'un Web multilingue soit souhaitable sur le plan moral et éthique, un tel idéal ne suffit pas pour en faire une réalité dépassant les limites actuelles. De même que l'utilisateur non anglophone peut maintenant avoir accès à la technologie 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 Internet doivent être préparés à faire face à 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!"

ML: "Quel a été le bénéfice de l'utilisation du Web dans votre activité?"

BK: "Le principal service que nous offrons est l'enseignement des langues par le biais du Web. Notre organisme est dans la position unique d'en être venu à exister du fait d'Internet!"

ML: "Comment voyez-vous l'évolution vers un Internet multilingue?"

BK: "Comme l'existence de notre organisme est liée à l'importance attachée aux langues, je pense que son avenir sera excitant et stimulant. Mais il est impossible de pratiquer l'autosuffisance à l'égard de nos réussites et de nos réalisations. 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 plus encore que par le passé, la coopération est aussi indispensable que la concurrence.

Les germes d'une coopération par le biais d'Internet existent déjà. Notre projet NetGlos a dépendu du bon vouloir de traducteurs volontaires de nombreux pays: Canada, Etats-Unis, Autriche, Norvège, Belgique, Israël, Portugal, Russie, Grèce, Brésil, Nouvelle-Zélande, etc. Je pense que 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."

#### 3.4. Bases de données textuelles

Des bases de données textuelles sont accessibles par abonnement payant, par exemple FRANTEXT et l'ARTFL Project pour la langue française.

FRANTEXT, présent sur le Web depuis début 1995, est l'oeuvre de l'Institut national de la langue française (INaLF), une branche du CNRS (Centre national de la recherche scientifique, France). La base comprend, en mode interactif, 180 millions de mots-occurrences résultant du traitement informatique d'une collection représentative de 3.500 unités textuelles en arts, sciences et techniques couvrant cinq siècles (16e-20e siècles). Début 1998, 82 centres de recherche et bibliothèques universitaires d'Europe, d'Australie, du Japon et du Canada étaient abonnés, ce qui représentait 1.250 postes de travail ayant accès à la base. Le nombre de sessions d'interrogations était d'une cinquantaine par jour.

Dans son courrier électronique du 11 juin 1998, Arlette Attali indiquait les changements apportés par Internet dans sa vie professionnelle:

"Etant moi-même plus spécialement affectée au développement des bases textuelles à l'INaLF, j'ai été amenée à explorer les sites du Web qui proposaient des textes électroniques et à les 'tester'. Je me suis donc transformée en 'touriste textuelle' avec les bons et mauvais côtés de la chose. La tendance au zapping et au survol étant un danger permanent, il faut bien cibler ce que l'on cherche si l'on ne veut pas perdre son temps. La pratique du Web a totalement changé ma façon de travailler: mes recherches ne sont plus seulement livresques et donc d'accès limité, mais elles s'enrichissent de l'apport des textes électroniques accessibles sur Internet.

[A l'avenir je pense] contribuer à développer des outils linguistiques associés à la base FRANTEXT et à les faire connaître auprès des enseignants, des chercheurs, des étudiants et aussi des lycéens."

En janvier 1998, elle a mené une enquête auprès des utilisateurs de FRANTEXT pour mieux connaître ses utilisateurs. Les résultats de cette enquête sont disponibles en ligne.

L'ARTFL Project (ARTFL: American and French Research on the Treasury of the French Language - Recherche franco-américaine sur les trésors de la langue française) est un projet commun du Centre national de la recherche scientifique (CNRS, France) et de l'Université de Chicago (Illinois, USA), qui vise à constituer une base de données de 2.000 textes du 13e au 20e siècle concernant la littérature, la philosophie, les arts ou les sciences.

L'ARTFL travaille aussi à la version en ligne exhaustive de la première édition (1751-1772) de l'Encyclopédie ou Dictionnaire raisonné des sciences, des métiers et des arts de Diderot et d'Alembert. 72.000 articles écrits par plus de 140 collaborateurs (dont Voltaire, Rousseau, d'Alembert, Marmontel, d'Holbach, Turgot, etc.) ont fait de cette encyclopédie un monumental ouvrage de référence pour les arts et les sciences. Destinée à rassembler puis divulguer les connaissances de l'époque, elle porte la marque des courants intellectuels et sociaux du 18e siècle, et c'est grâce à elle qu'ont été propagées les idées du Siècle des Lumières.

Les chiffres sont éloquentes quant à l'ampleur du travail: l'Encyclopédie comprend 17 volumes de texte et 11 volumes de planches, 18.000 pages de texte et 20.736.912 mots. La base de données correspondant au premier volume est accessible en ligne à titre expérimental. La recherche peut être effectuée par mot, portion de texte, auteur ou catégorie, ou par la combinaison de ces critères entre eux. On dispose de renvois d'un article à l'autre, et des liens permettent d'aller d'une planche au texte, ou du texte au fac-similé des pages originales. L'automatisation complète des procédures de saisie a entraîné quelques erreurs typographiques et des erreurs d'identification qui seront corrigées plus tard. La recherche d'images par mot, portion de texte ou catégorie sera également possible à l'avenir.

L'ARTFL travaille aussi à un projet de base de données pour le Dictionnaire de l'Académie française, dont les différentes éditions se sont échelonnées entre 1694 et 1935. Ce projet inclut la saisie et l'édition du texte, ainsi que le développement d'un moteur de recherche spécifique. Les différentes éditions pourront être combinées dans une seule base de données qui permettra de consulter aussi bien une édition particulière que l'ensemble de celles-ci pour juger de l'évolution d'un terme. Pour le

moment, seules deux éditions, la première (1694) et la cinquième (1798), sont disponibles pour une recherche par mot. Une fonction de recherche en texte intégral est prévue par la suite.

Une rubrique présente une liste des autres projets de l'ARTFL, notamment la version image de l'édition de 1740 du Dictionnaire historique et critique de Philippe Bayle, le Roget's Thesaurus de 1911, le Webster's Revised Unabridged Dictionary de 1913, le Thresor de la langue française de Jean Nicot (1606), un projet multilingue sur La Bible comprenant La Bible française de Louis Segond (1910), etc.

De par la quantité d'oeuvres dactylographiées à cette intention, le Project Gutenberg est la plus ancienne et la plus grande cyberbibliothèque qui existe. Créée en 1971 par Michael Hart aux Etats-Unis, elle a pour but de mettre gratuitement le plus grand nombre possible de textes à la disposition du plus grand nombre possible de lecteurs, à raison d'environ 45 titres par mois. Ses objectifs pour 2001 sont un stock de 10.000 textes littéraires et une transmission de 1.000 milliards de textes électroniques, soit 10.000 livres numériques vers 100 millions de lecteurs.

Le projet débuta en 1971 quand on donna à Michael Hart un compte de 100 millions de dollars de "temps machine" au Materials Research Lab de l'Université d'Illinois (USA). Immédiatement après avoir reçu ce crédit, il décida de le consacrer à la recherche et au stockage des oeuvres conservées dans les bibliothèques. Il décida aussi de stocker des textes électroniques de la manière la plus simple possible, en format ASCII, avec des lettres capitales pour les termes en italique, gras ou soulignés, afin que ces textes puissent être lus quels que soient la machine et le logiciel utilisés.

Cinquante heures environ sont nécessaires pour sélectionner, dactylographier, corriger et mettre en page un texte électronique. La dactylographie des textes est l'oeuvre de volontaires. Un ouvrage de taille moyenne - par exemple un roman de Stendhal ou de Jules Verne - est composé de deux fichiers ASCII.

Le Project Gutenberg inclut trois grands secteurs: la littérature de divertissement (Light Literature), comme Alice au pays des merveilles, Peter Pan ou les Fables d'Esopé, la littérature "sérieuse" (Heavy Literature) comme La Bible, les oeuvres de Shakespeare ou Moby Dick, et la littérature de référence (Reference Literature), composée d'encyclopédies et de dictionnaires, par exemple le Thesaurus de Roget.

Sur le site web, Michael Hart explique que la collection de littérature de divertissement est destinée à amener devant l'écran aussi bien un enfant d'âge pré-scolaire qu'une personne du troisième âge. Des enfants ou des grand-parents vont rechercher le texte électronique de Peter Pan après avoir vu Hook au cinéma, ou bien ils lisent Alice au pays des merveilles après l'avoir regardé à la télévision. Pratiquement tous les épisodes de Star Trek ont mentionné des livres qui ont leur correspondant électronique dans le Project Gutenberg (Moby Dick, Peter Pan...). L'objectif est que les gens puissent retrouver des citations qu'ils ont entendues dans des conversations, des films, des musiques, d'autres livres, et ce à l'aide d'une bibliothèque contenant tous ces éléments dans un format facile pour la recherche.

En juillet 1997, le Project Gutenberg fêtait son vingt-sixième anniversaire avec la mise en ligne des Merry Adventures of Robin Hood de Howard Pyle. En septembre 1997, il fêtait son millièm texte électronique avec la version anglaise de la Divine comédie de Dante. Dans sa lettre d'information d'octobre 1997, Michael Hart annonçait son intention de compléter la collection d'Oscar Wilde, de "séparer" les oeuvres complètes de Shakespeare en fichiers individuels pour chaque oeuvre, et de mettre en ligne des ouvrages non anglophones.

Outre l'anglais, on trouve quelques oeuvres en allemand, espagnol, français, italien et latin, mais elles ne sont pas encore légion. En janvier 1998, si on lançait une recherche sur les ouvrages disponibles en langue française, on trouvait neuf titres, dont six romans de Stendhal (L'Abbesse de Castro, La Chartreuse de Parme, La Duchesse de Palliano, Le Rouge et le Noir, Les Cenci, Vittoria Accorambani), deux romans de Jules Verne (De la terre à la lune et Le tour du monde en 80 jours) et French Cave Paintings, un ouvrage sur les peintures préhistoriques. A part ce dernier ouvrage, disponible depuis 1995, tous ces ouvrages n'ont été intégrés à la bibliothèque que début 1997. Si aucun titre de Stendhal n'était disponible en anglais, il existait trois oeuvres de Jules Verne dans cete langue: 20,000 Leagues Under the Sea (disponible depuis septembre 1994), Around the World in 80 Days (disponible depuis janvier 1994) et From the Earth to the Moon (disponible depuis septembre 1993).

Début septembre 1998, le nombre de titres d'ouvrages en langue française était monté à onze, avec Cyrano de Bergerac, d'Edmond Rostand, disponible depuis mars 1998, et La Révolution française, de Thomas Carlyle, disponible depuis mai 1998.

La lettre d'information d'octobre 1997 annonçait le développement des collections étrangères du Project Gutenberg. Dans celle de mars 1998, Michael Hart indiquait que les volontaires du projet

préparaient maintenant des textes électroniques en allemand, espagnol, français et portugais. Ces prochains mois, il espérait aussi recevoir des textes dans les langues suivantes: arabe, catalan, chinois, coréen, danois, espéranto, grec, hébreu, hollandais, hongrois, italien, japonais, latin, lituanien, polonais, roumain, russe, slovaque et slovène.

### 3.5. Bases de données terminologiques

La consultation gratuite sur le Web de bases terminologiques est très appréciée des linguistes. Voici par exemple quatre bases terminologiques gérées par des organisations internationales: Eurodicautom, géré par le Service de traduction de la Commission européenne, ILOTERM, géré par l'Organisation internationale du Travail (OIT), TERMITE (Base de données terminologique des Télécommunications de l'UIT), géré par l'Union internationale des télécommunications (UIT) et WHOTERM (WHO Terminology Information System), géré par l'Organisation mondiale de la santé (OMS).

Eurodicautom est la base de terminologie multilingue préparée par le Service de traduction de la Commission européenne. Mise en place à l'origine pour aider les traducteurs de la Commission, elle est consultée aujourd'hui par un nombre croissant de fonctionnaires de l'Union européenne autres que des traducteurs, ainsi que par des professionnels des langues dans le monde entier. Constamment mise à jour, cette base est disponible dans douze langues: allemand, anglais, danois, espagnol, finnois, grec, hollandais, italien, latin, portugais et suédois, et elle couvre de nombreux sujets.

ILOTERM est une base de données terminologique quadrilingue (allemand, anglais, espagnol et français) qui est gérée par l'Unité de terminologie et de références du Service des documents officiels (OFFDOC) de l'Organisation internationale du Travail (OIT). Comme l'indique le site web,

"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."

Géré par la section de traduction de l'Union internationale des télécommunications (UIT), TERMITE (Base de données terminologique des Télécommunications de l'UIT) est également quadrilingue (anglais, espagnol, français et russe).

"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 quelques 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 éditeurs techniques. TERMITE est surtout visité par les traducteurs internes mais aussi par des utilisateurs externes, travaillant dans le domaine des télécommunications."

Géré par l'Organisation mondiale de la santé (OMS), le système d'information terminologique WHOTERM (WHO Terminology Information System) est trilingue (anglais, espagnol et français). Il est constitué à partir des expressions et concepts des documents de l'OMS afin d'"améliorer la rigueur et la cohérence des textes rédigés, préparés ou traduits. Il 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."

## 4. TRADUCTION

[Dans ce chapitre:]

[4.1. Services de traduction / 4.2. Traduction automatique / 4.3. Traduction assistée par ordinateur]

### 4.1. Services de traduction

Créé par Vorontsoff, Wesseling & Partners (Amsterdam, Pays-Bas), Aquarius est un répertoire de traducteurs et interprètes incluant 6.100 traducteurs, 800 sociétés de traduction, 91 domaines d'expertise et 369 combinaisons de langues. Ce site non commercial permet de localiser et de contacter les meilleurs traducteurs directement, sans intermédiaire ni agence. La recherche est possible par lieu, combinaison de langues et spécialité.

Fondé by Bill Dunlap, Euro-Marketing Associates propose Global Reach, une méthode permettant aux sociétés d'étendre leur présence sur Internet dans un contexte international, ce qui comprend la traduction de leur site web dans d'autres langues, le promotion de ce site et l'utilisation de bandeaux publicitaires nationaux pour augmenter la consultation locale. Bill Dunlap précise:

"Promouvoir votre site est aussi important que de le créer, sinon plus. Vous devez être préparé à utiliser au moins autant de temps et d'argent à promouvoir votre site que vous en avez passé à l'origine à le créer. Le programme Global Reach vous permettra de promouvoir votre site dans des pays non anglophones, afin d'atteindre une clientèle plus large... et davantage de ventes. Il existe de nombreuses bonnes raisons pour considérer sérieusement le marché international. Global Reach est pour vous le moyen d'étendre votre site web à de nombreux pays, de le présenter à des visiteurs en ligne dans leur propre langue, et d'atteindre les marchés en ligne de ces pays."

Dans son courrier électronique du 11 décembre 1998, il expliquait aussi comment il en était venu à intégrer Internet dans sa vie professionnelle:

"Depuis 1981, début de ma vie 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 ici, 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. Récemment, lors de l'Internet World à New York, j'ai parlé du cybercommerce européen et de la manière d'utiliser un site web pour toucher les différents marchés d'Europe."

#### 4.2. Traduction automatique

La traduction automatique (TA) analyse le texte dans la langue-source et génère automatiquement le texte correspondant dans la langue-cible. L'être humain n'intervient pas au cours du processus, contrairement à la traduction assistée par ordinateur, qui exige une certaine interaction entre l'homme et la machine.

SYSTRAN, société spécialisée dans les logiciels de traduction, explique sur son site web:

"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."

La European Association for Machine Translation (EAMT) (Association européenne pour la traduction automatique) donne la définition suivante:

"La traduction automatique (TA) 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'organisations de traduction professionnelles."

Voici l'historique donné sur le site de Globalink, société spécialisée dans les logiciels et services de traduction :

"Dès leurs débuts, la traduction automatique et le traitement de la langue naturelle ont progressé de pair avec l'évolution de l'informatique quantitative. Le développement des premiers ordinateurs programmables pendant la Seconde guerre mondiale a été mené et accéléré par les premiers efforts cryptographiques pour tenter de fissurer les codes secrets allemands et autres codes de guerre. Suite à la guerre, la traduction et l'analyse du texte en langue naturelle procura une base de travail au secteur émergent de la théorie de l'information.

Pendant les années 50, la recherche sur la traduction automatique prit forme dans le sens de traduction littérale, ou traduction mot à mot, sans utiliser de règles linguistiques.

Le projet russe débuté à l'Université de Georgetown au début des années 50 représentait la première tentative systématique pour créer un système de traduction automatique utilisable. Tout au long des

années 50 et au début des années 60, un certain nombre de recherches universitaires et recherches financées par les gouvernements furent menées aux Etats-Unis et en Europe. Au même moment, les progrès rapides dans le domaine de la linguistique théorique culminaient en 1965 avec la publication de *Aspects of the Theory and Syntax* (Aspects de la théorie et de la syntaxe) de Noam Chomsky, et transformaient radicalement la structure permettant de discuter et comprendre la phonologie, la morphologie, la syntaxe et la sémantique du langage humain.

En 1966, le rapport ALPAC du gouvernement des Etats-Unis faisait une estimation prématurément négative de la valeur et des perspectives des systèmes d'application pratique de la traduction automatique, mettant ainsi fin au financement et à l'expérimentation dans ce domaine pour la décennie suivante. Ce fut seulement à la fin des années 70, avec le développement de la technologie de l'informatique et des langues, que des tentatives sérieuses furent à nouveau entreprises. Cette période d'intérêt renouvelé vit aussi le développement du modèle de transfert de la traduction automatique et l'émergence des premiers systèmes commerciaux de traduction automatique.

Des entreprises commerciales comme SYSTRAN et METAL commençaient à prouver que la traduction automatique était viable et utile. Parallèlement à la mise sur le marché de produits et services de traduction automatique, ces systèmes liés à un processeur central mettaient aussi en lumière de nombreux problèmes. Des coûts élevés de développement, une lexicographie émanant d'un travail intensif et son implémentation linguistique, le lent développement de nouvelles combinaisons de langues, l'inaccessibilité pour l'utilisateur moyen et l'incapacité d'accéder aisément à de nouveaux stades de développement sont les caractéristiques de ces systèmes de la seconde génération."

Un certain nombre de sociétés sont spécialisées dans le développement de la traduction automatique, par exemple Lernout & Hauspie, Globalink, Logos ou SYSTRAN.

Basé à Ypres (Belgique) et Burlington (Massachusetts, USA), Lernout & Hauspie (L&H), leader international dans ce domaine, développe une technologie avancée du langage dans diverses applications et produits commerciaux. La société offre quatre technologies de base: la reconnaissance automatique de la langue, la compression numérique de la parole, le passage du texte à la parole et le passage du texte au texte. Les produits émanant des trois premières technologies sont vendus aux grandes sociétés des industries suivantes: télécommunications, informatique, multimédias, électronique grand public et électronique automobile. Les services de traduction (passage du texte au texte) sont à destination des sociétés de technologie de l'information, des marchés verticaux et des marchés d'automatisation.

Le Machine Translation Group (Groupe de traduction automatique) de Lernout & Hauspie comprend des entreprises qui développent, produisent et vendent des systèmes de traduction automatique hautement sophistiqués: L&H Language Technology, AppTek, AILogic, NeocorTech et Globalink. Chaque entreprise est un leader international dans sa propre partie.

Fondé en 1990, Globalink est une société américaine spécialisée dans les logiciels et services de traduction. Elle offre des solutions sur mesure à partir d'une gamme de logiciels, options en ligne et services de traduction professionnelle. La société diffuse ses logiciels de traduction en allemand, anglais, espagnol, français, italien et portugais, et propose des solutions aux problèmes de traduction à tous niveaux: particuliers, petites sociétés, multinationales et gouvernements, que ce soit pour un produit individuel donnant une traduction préliminaire rapide ou un système complet permettant de gérer des traductions de documents professionnels.

Le site web donne les informations suivantes:

"Avec les logiciels d'application de Globalink, l'ordinateur utilise trois ensembles de données : le texte à traiter, le programme de traduction et un dictionnaire de mots et d'expressions dans la langue-source, ainsi que des informations sur les concepts évoqués par le dictionnaire et les règles applicables à la phrase: règles de syntaxe et de grammaire, y compris des algorithmes gouvernant la conjugaison des verbes, l'adaptation de la syntaxe, les accords de genre et de nombre et la mise en ordre des mots.

Une fois que l'utilisateur a sélectionné le texte et lancé le processus de traduction, le programme commence à comparer les mots du texte à traiter avec ceux qui sont stockés dans le dictionnaire. Une fois l'adéquation trouvée, l'application prépare une notice complète qui inclut des informations sur les significations possibles du mot et, d'après le contexte, ses relations avec les autres mots dans la même phrase. Le temps requis pour la traduction dépend de la longueur du texte. Un document de trois pages et 750 mots demande un traitement de trois minutes environ pour une première traduction."

Randy Hobler est directeur de marketing pour les produits et services Internet de Globalink. Dans son courrier électronique du 3 septembre 1998, il écrivait:

"En 1998, 85 % du contenu du Web est en anglais, et ce chiffre est à la baisse. Il y a non seulement plus de sites web et d'internautes non anglophones, mais aussi une localisation plus grande de sites de sociétés et d'organismes, et un usage accru de la traduction automatique pour traduire des sites web à partir ou vers d'autres langues.

Comme 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 des langues', tous ces internautes où qu'ils soient qu'on peut regrouper selon leur langue maternelle. 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 Etats-Unis, ou encore ceux qui parlent espagnol au Maroc.

Concernant la transparence de la langue, nous arrivons rapidement au point où une traduction très fidèle du texte et de la parole sera si commune qu'elle pourra faire partie des plate-formes ou même des puces. A ce point, quand le développement d'Internet aura atteint sa vitesse de croisière, que la fidélité de la traduction atteindra plus de 98% et que les différentes combinaisons de langues possibles auront couvert la grande majorité du marché, la transparence de la langue (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 sociétés. La lettre O réalisée avec le pouce et l'index signifie 'OK' aux Etats-Unis alors qu'en Argentine c'est un geste obscène.

Quand se produira l'inévitable développement de la vidéoconférence multilingue multimédias, il sera nécessaire de corriger visuellement les gestes. Le Media Lab du MIT [MIT: Massachusetts Institute of Technology], Microsoft et bien d'autres travaillent à la reconnaissance informatique des expressions faciales, l'identification des caractéristiques biométriques par le biais du visage, etc. Il ne servira à rien à un homme d'affaires américain de faire une excellente présentation à un Argentin lors d'une vidéoconférence multilingue sur le Web, avec son discours traduit dans un espagnol argentin parfait, s'il fait en même temps le geste O avec le pouce et l'index. Les ordinateurs pourront intercepter ces types de messages et les corriger visuellement.

Les cultures diffèrent de milliers de façons, et la plupart d'entre elles peuvent être modifiées par voie informatique lorsqu'on passe de l'une 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 mesures anglophone, etc., etc. Les sociétés 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 ceux-ci seront largement répandus, ils contribueront réellement à une meilleure compréhension à l'échelle internationale."

Basée aux Etats-Unis, au Canada et en Europe, Logos est une société internationale spécialisée dans la traduction automatique depuis 25 ans. Elle procure différents outils de traduction, des systèmes de traduction automatique et des services de soutien.

SYSTRAN (acronyme de System Translation) est également une société spécialisée dans les logiciels de traduction automatique. Son siège est situé à Soisy-sous-Montmorency (France). C'est sa succursale, située à La Jolla (Californie), qui assure les ventes, le marketing et le développement des logiciels. Une des réalisations de la société est AltaVista Translation, un service de traduction automatique de pages web de l'anglais vers les langues suivantes : allemand, français, espagnol, italien et portugais, et vice versa. Ce service a été mis en place en décembre 1997 à la demande d'AltaVista, moteur de recherche utilisé par douze millions d'internautes, suite au problème des langues devenu sensible sur Internet.

Basée à Montréal (Québec), Alis Technologies développe et commercialise des solutions et services de traitement linguistique au moyen de logiciels, périphériques et solutions de traduction qui transforment des systèmes informatiques unilingues en outils multilingues.

Une autre réalisation intéressante est SPANAM/ENGSPAN, un système de traduction automatique développé par les linguistes computationnels, les traducteurs et le programmeur systèmes de l'Organisation panaméricaine de la santé (PAHO) (Bureau régional de l'Organisation mondiale de la santé pour les Amériques, situé à Washington, D.C., USA). Depuis 1980, le service de traduction utilise SPANAM (de l'espagnol vers l'anglais) et ENGSPAN (de l'anglais vers l'espagnol), ce qui lui a permis de traiter plus de 25 millions de mots dans les deux langues de travail de la PAHO. Le personnel et les traducteurs extérieurs post-éditent ensuite l'information brute avec un gain de productivité de 30 à 50%. Le système est installé sur le réseau local du siège de l'organisation et dans un certain nombre de bureaux régionaux pour pouvoir être utilisé par le personnel des services techniques et administratifs. Il est également diffusé auprès d'organismes publics et d'organismes à but non lucratif aux Etats-Unis,

en Amérique latine et en Espagne.

Des associations contribuent au développement de la traduction automatique.

L'Association for Computational Linguistics (ACL) (Association pour la linguistique computationnelle) est le principal organisme international à la fois scientifique et professionnel rassemblant ceux qui travaillent sur les problèmes de la langue naturelle et de la computation. Publiée par la MIT Press, la revue trimestrielle de l'ACL, *Computational Linguistics* (ISSN 0891-2017) est un forum de premier plan dans le domaine de la linguistique computationnelle et du traitement de la langue naturelle. Cette revue est complétée par la lettre d'information *The Finite String*. La branche européenne de l'ACL est l'European Chapter of the Association of Computational Linguistics (EACL).

L'International Association for Machine Translation (IAMT) (Association internationale pour la traduction automatique) regroupe trois associations couvrant les Amériques, l'Europe et l'Asie/Pacifique: l'Association for Machine Translation in the Americas (AMTA), la European Association for Machine Translation (EAMT) et l'Asia-Pacific Association for Machine Translation (AAMT).

L'Association for Machine Translation in the Americas (AMTA) (Association pour la traduction automatique dans les Amériques) est à la disposition de tous ceux s'intéressent à la traduction automatique en Amérique latine, au Canada et aux Etats-Unis. Ses membres comprennent des demandeurs de traductions, des développeurs de systèmes commerciaux, des chercheurs, des sponsors et des personnes impliquées dans la science de la traduction automatique et sa diffusion.

Basée à Genève (Suisse), l'European Association for Machine Translation (EAMT) (Association européenne pour la traduction automatique) est également une organisation à la disposition de ceux qui s'intéressent à la traduction automatique et aux outils de traduction, y compris les utilisateurs de ces techniques, les développeurs et les chercheurs.

Créée en 1991, la Japan Association for Machine Translation, devenue ensuite l'Asia-Pacific Association for Machine Translation (AAMT) (Association de l'Asie et du Pacifique pour la traduction automatique), comprend des chercheurs, des fabricants et des utilisateurs de systèmes de traduction automatique. L'association participe au développement des technologies de traduction automatique, ainsi qu'à leur amélioration, leur enseignement et leur diffusion.

Dans *Web embraces language translation* (La traduction des langues sur le Web), un article de ZDNN (ZD Network News) paru le 21 juillet 1998, Martha L. Stone expliquait:

"Parmi les nouveaux produits d'un secteur de traduction représentant 10 milliards de dollars [60 milliards de FF], on trouve les traducteurs instantanés de sites web, de groupes de discussion, de courrier électronique et d'intranets d'entreprise.

Les principales sociétés de traduction se mobilisent pour saisir les opportunités du marché. Voici quelques exemples.

SYSTRAN s'est associé avec AltaVista pour produire [babelfish.altavista.digital.com](http://babelfish.altavista.digital.com), avec 500 à 600 mille visiteurs quotidiens et environ un million de traductions par jour, traductions qui vont des recettes à des pages web complètes.

15.000 sites environ ont un lien vers babelfish, qui peut traduire [de l'anglais] vers le français, l'italien, l'allemand, l'espagnol et le portugais, et vice versa. Le japonais est prévu pour bientôt.

"Cette popularité est simple. Avec Internet, on peut maintenant utiliser l'information provenant des Etats-Unis. Tout ceci contribue à une demande en hausse", déclare de chez lui à Paris Dimitros Sabatakakis, directeur général de SYSTRAN.

Alis a mis au point le système de traduction du Los Angeles Times qui doit bientôt être lancé sur le site et qui proposera des traductions [de l'anglais] vers l'espagnol et le français, et plus tard le japonais. D'un clic de souris, une page web complète peut être traduite dans la langue désirée.

Globalink propose des logiciels, des systèmes de traduction de pages web, un service de messagerie électronique gratuit et des logiciels permettant de traduire le texte des groupes de discussion.

Cependant, alors que ces systèmes de traduction automatique deviennent populaires dans le monde entier, les directeurs des sociétés qui les développent admettent qu'ils ne peuvent répondre à toutes les situations.

Les porte-parole de Globalink, Alis et SYSTRAN utilisent des expressions comme "pas parfait" et "approximatif" quand ils décrivent la qualité des traductions, et précisent bien que les phrases soumises à la traduction doivent être simples, grammaticalement correctes et sans tournures



idiomatiques.

"Les progrès réalisés en traduction automatique répondent à la loi de Moore: la qualité double tous les dix-huit mois, déclare Vin Crosbie, un analyste de l'industrie du Web basé à Greenwich, dans le Connecticut [USA]. "Ce n'est pas parfait, mais certains de mes correspondants ne se rendent même pas compte que j'utilise un logiciel de traduction."

Ces traductions font souffrir la syntaxe et l'utilisation des mots à bon escient, parce que les bases de données-dictionnaires ne peuvent déchiffrer la différence entre les homonymes [...].

"La traduction humaine coûterait entre 50 et 60 dollars [300 à 360 FF] par page web, ou environ 20 cents [1,15 FF] par mot", explique Sabatakis, directeur de SYSTRAN.

Alors que cette dernière solution peut convenir pour les pages 'statiques' d'information sur l'entreprise, la traduction automatique, elle, est gratuite sur le Web, et le logiciel coûte souvent moins de 100 dollars [600 FF], selon le nombre de langues disponibles pour traduction et les caractéristiques propres au logiciel."

#### 4.3. Traduction assistée par ordinateur

Contrairement à la traduction automatique (TA) qui, sans intervention humaine, analyse le texte dans la langue-source et génère automatiquement le texte correspondant dans la langue-cible, la traduction assistée par ordinateur (TAO) exige une certaine interaction entre l'homme et la machine.

Le Bureau des services linguistiques de l'Organisation mondiale de la santé (OMS) (Genève, Suisse) travaille dans les six langues officielles de l'organisation: anglais, arabe, chinois, espagnol, français et russe.

Des expériences de traduction automatique ont été tentées à plusieurs reprises, mais les traductions obtenues demandaient un travail de révision trop important, si bien que, au stade actuel de son développement et compte tenu du type de documents à traduire, cette technologie n'a pas été jugée suffisamment rentable.

L'Unité de Traduction assistée par ordinateur et de Terminologie (CTT) explore donc les possibilités techniques qu'offrent les systèmes les plus récents de traduction assistée par ordinateur, reposant sur la notion de "mémoire de traduction". Comme il est expliqué sur le site web,

"ces systèmes permettent au traducteur d'avoir immédiatement accès au patrimoine du 'déjà traduit' dans lequel il peut puiser, quitte à rejeter ou modifier les solutions retenues par ses prédécesseurs, son choix définitif venant ensuite enrichir la mémoire. Ainsi, en archivant la production quotidienne, le traducteur aurait vite à sa disposition une 'mémoire' colossale de solutions toutes faites à un nombre important de problèmes de traduction."

Le CTT a entrepris l'évaluation de plusieurs applications dans les domaines suivants: archivage électronique et recherche en texte intégral, alignement de textes bilingues et multilingues, traduction assistée par ordinateur, gestion de mémoires de traduction et de bases de données terminologiques, et reconnaissance vocale.

Contrairement aux prévisions optimistes d'il y a cinquante ans annonçant l'apparition imminente de la machine à traduire universelle, les systèmes de traduction automatique ne produisent toujours pas de traductions de bonne qualité. Pourquoi? Pierre Isabelle et Patrick Andries, du Laboratoire de recherche appliquée en linguistique informatique (RALI) (Montréal, Québec) expliquent ce échec dans La traduction automatique, 50 ans après, un article publié dans les Dossiers du cyberquotidien Multimédium:

"L'objectif ultime de construire une machine capable de rivaliser avec le traducteur humain n'a cessé de fuir par devant les lentes avancées de la recherche. Les approches traditionnelles à base de règles ont conduit à des systèmes qui tendent à s'effondrer sous leur propre poids bien avant de s'élever au-dessus des nuages de l'ambiguïté sémantique. Les approches récentes à base de gros ensembles de textes, appelés corpus - qu'elles soient fondées sur les méthodes statistiques ou les méthodes analogiques - promettent bien de réduire la quantité de travail manuel requise pour construire un système de TA [traduction automatique], mais il est moins sûr qu'elles promettent des améliorations substantielles de la qualité des traductions machine."

Reprenant les idées de Yehochua Bar-Hillel exprimées dans The State of Machine Translation (L'état de la traduction automatique), article publié en 1951, Pierre Isabelle et Patrick Andries définissent trois stratégies d'application de

la traduction automatique:

a) une aide pour "balayer" la production écrite et fournir des traductions approximatives,

b) des situations de "sous-langues naturelles simples", comme l'implantation réussie en 1977 du système METEO qui traduit les prévisions météorologiques du ministère de l'Environnement canadien,

c) pour de très bonnes traductions de textes complexes, le couplage de l'humain et de la machine avant, pendant et après le processus de traduction automatique, couplage qui n'est pas forcément économique comparé à la traduction traditionnelle.

Les auteurs penchent plus pour "un poste de travail pour le traducteur humain" que pour un "traducteur robot":

"Les recherches récentes sur les méthodes probabilistes ont en effet permis de démontrer qu'il était possible de modéliser d'une manière extrêmement efficace certains aspects simples du rapport traductionnel entre deux textes. Par exemple, on a mis au point des méthodes qui permettent de calculer le bon "appariement" entre les phrases d'un texte et de sa traduction, c'est-à-dire d'identifier à quelle(s) phrase(s) du texte d'origine correspond chaque phrase de la traduction. Appliquées à grande échelle, ces techniques permettent de constituer, à partir des archives d'un service de traduction, un mémoire de traduction qui permettra souvent de recycler des fragments de traduction antérieures. Des systèmes de ce genre ont déjà commencé à apparaître sur le marché (Translation Manager II de IBM, Translator's Workbench de Trados, TransSearch du RALI, etc.).

Les recherches les plus récentes se concentrent sur des modèles capables d'établir automatiquement les correspondances à un niveau plus fin que celui de la phrase: syntagmes et mots. Les résultats obtenus laissent entrevoir toute une famille de nouveaux outils pour le traducteur humain, dont les aides au dépouillement terminologique, les aides à la dictée et à la frappe des traductions ainsi que les détecteurs de fautes de traduction."

## 5. RECHERCHE

[Dans ce chapitre:]

[5.1. Traduction automatique et recherche / 5.2. Linguistique computationnelle / 5.3. Ingénierie du langage / 5.4. Internationalisation et localisation]

### 5.1. Traduction automatique et recherche

Au sein du Laboratoire CLIPS (Communication langagière et interaction personne-système) de la Fédération IMAG (France), le Groupe d'étude pour la traduction automatique (GETA) est une équipe pluridisciplinaire formée d'informaticiens et de linguistes. Les thèmes de recherche du GETA concernent tous les aspects théoriques, méthodologiques et pratiques de la traduction assistée par ordinateur (TAO), et plus généralement de l'informatique multilingue. Le GETA est issu du CETA (1961-1971), laboratoire pionnier de la traduction automatique en France.

Actuellement, le GETA reste actif en TAO du réviseur, mais réoriente sa recherche, depuis 1988, vers la TAO individuelle, qui comporte deux volets, la TAO du traducteur et celle du rédacteur. Les définitions de ces deux TAO sont données sur le site:

"La TAO du traducteur consiste à offrir des outils de bureautique linguistique à des traducteurs (professionnels ou occasionnels). C'est l'homme qui traduit. Dans ce domaine, nous travaillons en coopération avec d'autres groupes de recherche, qui apportent des données ou outils linguistiques (lexiques, lemmatiseurs), et nous nous intéressons aux problèmes informatiques liés à l'intégration de ces éléments sous une forme utilisable par des traducteurs occasionnels, désireux de les employer depuis leurs applications favorites. Nous avons récemment élaboré et proposé en collaboration avec SITE-Eurolang, le projet Montaigne, qui vise à mettre à disposition de la communauté scientifique le logiciel Eurolang-Optimizer, via Internet, et à l'utiliser pour constituer des grosses bases terminologiques pouvant ensuite aussi alimenter des systèmes automatiques.

La TAO du rédacteur est l'objectif principal des travaux en cours, regroupés dans le projet LIDIA. L'idée de base est d'offrir à un rédacteur unilingue la possibilité de rédiger dans sa langue, et, au prix d'un dialogue de standardisation et de désambiguïsation (qu'il conviendra de rendre le moins lourd et le plus convivial possible), d'être traduit dans plusieurs langues, sans révision ou avec une révision minimale. Il s'agit donc de TAO fondée sur le dialogue (DBMT, pour dialogue-based machine

translation) et de prédiction indirecte, mais c'est bien la machine qui traduit. Une première maquette, LIDIA-1, partant du français et allant vers l'allemand, le russe et l'anglais, a été spécifiée et réalisée au cours des trois dernières années.

Le GETA participe aussi au projet de l'Universal Networking Language (UNL) sous l'égide de l'Université des Nations Unies (ONU). Voici la présentation du projet préparée en septembre 1998 par Christian Boitet, directeur du GETA:

"Il s'agit non de TAO habituelle, mais de communication et recherche d'information multilingue. 14 groupes ont commencé le travail sur 12 langues (plus 2 annexes) depuis début 97. L'idée est:

- développer un standard, dit UNL, qui serait le HTML du contenu linguistique.
- pour chaque langue, développer un générateur (dit 'déconvertisseur') accessible sur un ou plusieurs serveurs, et un 'enconvertisseur'.

L'ONU (Université des Nations Unies, Tokyo) finance 50% du coût. D'après notre évaluation sur la première année, c'est plutôt 30 à 35%, car le travail (linguistique et informatique) est énorme, et le projet passionnant: les permanents des laboratoires s'y investissent plus que prévu. [...]

La déconversion tourne pour le japonais, le chinois, l'anglais, le portugais, l'indonésien, et commence à tourner pour le français, l'allemand, le russe, l'italien, l'espagnol, l'hindi, l'arabe, et le mongol.

Chaque langue a une base lexicale de 30.000 à 120.000 liens UW [universal word]—lexème.

L'enconversion n'est pas (si on veut de la qualité pour du tout venant) une analyse classique. C'est une méthode de fabrication de graphes UNL qui suppose une bonne part d'interaction, avec plusieurs possibilités :

- analyse classique multiple suivie d'une désambiguïsation interactive en langue source,
- entrée sous langage contrôlé,
- encore plus séduisant (et encore pas clair, au niveau recherche pour l'instant), entrée directe via une interface graphique reliée à la base lexicale et à la base de connaissances.

Applications possibles :

- courriel multilingue
- informations multilingues
- dictionnaires actifs pour la lecture de langues étrangères sur le Web
- et bien sûr TA [traduction automatique] de mauvaise qualité (ce qu'on trouve actuellement, mais pour tous les couples à cause de l'architecture à pivot) pour le surf web et la veille.

On travaille actuellement sur les informations sportives sur le Web, surtout sur le foot. On construit une base de documents, où chaque fichier est structuré (à la HTML) et contient, pour chaque énoncé, l'énoncé original, sa structure UNL, et autant de traductions qu'on en a obtenu. Un tel document peut être recherché dans une base en traduisant la question en UNL, puis affiché (le UNL viewer existe depuis un an) dans autant de fenêtres d'un browser Web que de langues sélectionnées.

Perspectives:

Le projet a un problème de volume: grande surface, pas assez d'épaisseur. Il faudrait 3 à 5 fois plus de monde partout pour que ça avance assez vite (pour que Microsoft et d'autres ne finissent pas par tout reprendre et revendre, alors qu'on vise une utilisation ouverte, du type de ce qu'on fait avec les serveurs et clients Web). Les subventions des sociétés japonaises à l'ONU pour ce projet (et d'autres) se tarissent à cause de la crise japonaise. Le groupe central est beaucoup trop petit (4 personnes qui font le logiciel, le japonais, l'anglais, l'administration, c'est peu même avec de la sous-traitance).

De plus, le plan général est d'ouvrir aux autres langues de l'ONU en 2000. Il faudrait arriver à un état satisfaisant pour les 13 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.

Dans l'état actuel des choses, je pense que l'élan initial a été donné, mais que la première phase (d'ici 2000) risque de retomber comme un soufflé si on ne consolide pas très vite le projet, dans chaque pays participant.

Et donc:

L'ONU cherche comment monter un soutien puissant à la mesure de cette ambition. Je pense que, pour la francophonie par exemple, il faudrait un groupe d'une dizaine de personnes ne se consacrant qu'à ce projet pendant au moins 10 ans, plus des stagiaires et des collaborateurs sur le réseau, bénévoles ou intéressés par la mise à disposition gratuite de ressources et d'outils."

Voici quelques groupes de recherche anglophones.

Créé à la fin des années 80, le CL/MT Research Group (Computational Linguistics (CL) and Machine Translation (MT) Group - Groupe de linguistique computationnelle et de traduction automatique) est un groupe de recherche du Département des langues et de linguistique de l'Université d'Essex (Royaume-Uni).

Fondé en 1986, le Center for Machine Translation (CMT) (Centre pour la traduction automatique) est un centre de recherche inclus dans le nouvel Institut des technologies des langues de l'Université Carnegie Mellon (Pittsburgh, Pennsylvanie, USA). Il est spécialisé dans les technologies de traitement de la langue naturelle, particulièrement la traduction automatique multilingue de haute qualité.

Le Computing Research Laboratory (CRL) (Laboratoire de recherche informatique) à la New Mexico State University (NMSU) (Université d'Etat du Nouveau Mexique - Etats-Unis) est un centre de recherche à but non lucratif qui se consacre à la recherche de base et au développement de logiciels dans les applications informatiques avancées du traitement de la langue naturelle, de l'intelligence artificielle et de la conception d'interfaces graphiques pour les utilisateurs. Les applications développées à partir de cette recherche de base incluent un ensemble de configurations de traduction automatique, extraction d'information, acquisition du savoir, enseignement intelligent, et systèmes de traduction pour poste de travail.

Géré par le Département de linguistique du Groupe de recherche sur la traduction de l'Université Brigham Young (Utah, USA), TTT.org (Translation, Theory and Technology) (Traduction, théorie et technologie) procure des informations sur la théorie et la technologie de la langue, particulièrement dans le domaine de la traduction. La technologie de la traduction inclut les outils de traduction pour poste de travail et la traduction automatique.

TTT.org s'intéresse aussi aux normes d'échanges de données permettant à divers outils de fonctionner en lien les uns avec les autres, ce qui permet l'intégration d'outils de la part de vendeurs multiples dans la chaîne multilingue de production de documents. Dans ce domaine, TTT.org est impliqué dans le développement de MARTIF (machine-readable terminology interchange format - format d'échange de terminologie lisible par la machine), un format permettant de faciliter l'interaction de données terminologiques entre les systèmes de gestion de terminologie. Ce format est le résultat de plusieurs années de collaboration internationale intense entre les terminologues et les experts en bases de données de plusieurs organisations, incluant les organismes universitaires, la Text Encoding Initiative (TEI) (Initiative pour le codage du texte) et la Localisation Industry Standards Association (LISA) (Association pour les normes de l'industrie de la localisation).

Le Natural Language Group (NLG) (Groupe de langue naturelle) de l'USC/ISI (University of Southern California/Information Sciences Institute - Université de la Californie du Sud/Institut des sciences de l'information) traite de plusieurs aspects du traitement de la langue naturelle: traduction automatique, résumé automatique de texte, accès multilingue aux verbes et gestion du texte, développement de taxonomies de concepts (ontologies), discours et génération de texte, élaboration de grands lexiques pour plusieurs langues, et communication multimédias.

Eduard Hovy, directeur du Natural Language Group, expliquait dans son courrier électronique du 27 août 1998:

"Le plan de votre étude me paraît très intéressant. Je me demande cependant où vous présentez les applications et fonctionnalités n'ayant pas trait à la traduction, comme la recherche documentaire et le résumé automatique de texte. On ne peut rien trouver sur le Web sans recherche documentaire, et tous les engins de recherche (AltaVista, Yahoo!, etc.) sont basés sur cette technologie. De même, bien que ceci soit plus récent, il y a des chances pour que beaucoup de gens utilisent bientôt des fonctions de résumé automatique pour condenser ou extraire le contenu principal d'un document long ou d'un ensemble de documents. [...]"

Dans ce contexte, le multilinguisme sur le Web est un autre facteur de complexité. Les gens écrivent dans leur propre langue pour diverses raisons: commodité, discrétion, communication locale, mais ceci ne signifie pas que d'autres personnes ne soient pas intéressées de lire ce qu'ils ont à dire! Ceci est particulièrement vrai pour les sociétés impliquées dans la veille technologique (disons, une société informatique qui souhaite connaître tous les articles de journaux et périodiques japonais relatifs à son activité) et des services de renseignements gouvernementaux (ceux qui procurent l'information la plus récente qui sera ensuite utilisée par les fonctionnaires pour décider de la politique, etc.). Un des principaux problèmes auquel ces services doivent faire face est la très grande quantité d'informations. Ils recrutent donc du personnel bilingue 'passif' qui peut scanner rapidement les textes afin de supprimer ce qui est sans intérêt avant de donner les documents significatifs à des traducteurs professionnels. Manifestement, une combinaison de résumé automatique de texte et de traduction automatique sera très utile dans ce cas; comme la traduction automatique est longue, on peut d'abord résumer le texte dans la langue étrangère, puis faire une traduction automatique rapide à partir du résultat obtenu, laissant à un être humain ou un classificateur de texte (type recherche documentaire) le soin de décider si on doit garder l'article ou le rejeter.

Pour ces raisons, durant ces cinq dernières années, le gouvernement des Etats-Unis a financé des recherches en traduction automatique, en résumé automatique de texte et en recherche documentaire, et il s'intéresse au lancement d'un nouveau programme de recherche en informatique documentaire multilingue. On sera ainsi capable d'ouvrir un navigateur tel que Netscape ou Explorer, entrer une demande en anglais, et obtenir la liste des textes dans toutes les langues. Ces textes seront regroupés par sous-catégorie avec un résumé pour chacun et une traduction pour les résumés étrangers, toutes choses qui seraient très utiles.

En consultant le MuST Multilingual Information Retrieval, Summarization, and Translation System (Système MuST de recherche documentaire, résumé et traduction multilingues), vous aurez une démonstration de notre version de ce programme de recherche, qui utilise l'anglais comme langue de l'utilisateur sur un ensemble d'environ 5.000 textes en anglais, japonais, arabe, espagnol et indonésien.

Entrez votre demande (par exemple, 'baby', ou ce que vous voulez) et appuyez sur la touche 'Retour'. Dans la fenêtre du milieu vous verrez les titres (ou bien les mots-clés, traduits). Sur la gauche vous verrez la langue de ces documents: 'Sp' pour espagnol, 'Id' pour indonésien, etc. Cliquez sur le numéro situé sur la partie gauche de chaque ligne pour voir le document dans la fenêtre du bas. Cliquez sur 'Summarize' pour obtenir le résumé. Cliquez sur 'Translate' pour obtenir la traduction (attention, les traductions en arabe et en japonais sont extrêmement lentes! Essayez plutôt l'indonésien pour une traduction rapide mot à mot).

Ce programme de démonstration n'est pas (encore) un produit. Nous avons de nombreuses recherches à mener pour améliorer la qualité de chaque étape. Mais ceci montre la direction dans laquelle nous allons."

## 5.2. Linguistique computationnelle

Le Laboratoire de recherche appliquée en linguistique informatique (RALI) (Montréal, Québec) réunit des informaticiens et des linguistes d'expérience dans le traitement automatique de la langue tant par des méthodes symboliques "classiques" que par de nouvelles méthodes probabilistes.

Un rapide historique du RALI est donné sur le site web: grâce au laboratoire Incognito fondé en 1983, le département d'informatique et de recherche opérationnelle (DIRO) de l'Université de Montréal avait acquis une stature de premier plan en matière de recherche en traitement automatique de la langue naturelle. En juin 1997, le DIRO a obtenu du ministère de l'Industrie du gouvernement canadien l'impartition du programme de recherche en traduction assistée par ordinateur (TAO) poursuivi depuis 1984 par le Centre d'innovation en technologie de l'information (CITI). C'est dans ce cadre qu'a été mis sur pied le RALI qui permet de mettre en valeur les résultats de ces recherches.

Les domaines de compétence du RALI sont les suivants: outils d'aide à la traduction, appariement automatique de textes, génération automatique de texte, réaccentuation automatique, recherche d'information aidée par des outils linguistiques, extraction d'information, identification de la langue et du codage, et transducteurs à états finis.

Dans le cadre du Projet TransX, le RALI élabore une nouvelle génération d'outils d'aide aux traducteurs (TransType, TransTalk, TransCheck et TransSearch). Ces outils sont tous fondés sur des modèles de traduction probabilistes qui calculent automatiquement les correspondances entre le texte produit par le traducteur et le texte en langue de départ.

"TransType accélère la saisie de la traduction en anticipant les choix du traducteur et, au besoin, en

les critiquant. L'outil propose ses choix en tenant compte à la fois du texte en langue de départ et de la traduction partielle déjà produite par le traducteur.

TransTalk effectue la transcription automatique d'une traduction dictée. Cet outil se sert d'un modèle de traduction probabiliste pour améliorer la performance du module de reconnaissance vocale.

TransCheck détecte automatiquement certaines erreurs de traduction en vérifiant que les correspondances entre les segments d'une ébauche de traduction et les segments du texte en langue de départ respectent les propriétés souhaitées d'une bonne traduction.

TransSearch permet au traducteur d'effectuer des recherches dans des bases de données de traductions pré-existantes pour y retrouver des solutions toutes faites à ses problèmes de traduction. Les bases de données requises nécessitent un appariement entre la traduction et le texte en langue de départ."

Financés par la société XEROX, les projets du Xerox Palo Alto Research Center (PARC) incluent deux projets relatifs aux langues: Inter-Language Unification (ILU) et Natural Language Theory and Technology (NLTT).

L'Inter-Language Unification (ILU) System (Système d'unification inter-langues) est un système d'interface-objet multi-langues. Les interfaces-objet fournis par l'ILU dissimulent les différences d'implémentation entre les différentes langues, les différents espaces d'adresse et les types de systèmes d'exploitation. ILU peut être utilisé pour construire des bibliothèques multilingues orientées vers l'objet ("bibliothèques de catégorie") avec des interfaces bien spécifiés indépendants des langues. Il peut être utilisé aussi pour réaliser des systèmes distribués, ou pour définir et établir les documents des interfaces entre les modules de programmes non distribués.

Le but de la Natural Language Theory and Technology (NLTT) (Théorie et technologie de la langue naturelle) est de développer des théories sur la façon dont l'information est codée dans la langue naturelle et dans la technologie pour organiser l'information vers et à partir des représentations en langue naturelle. Le but est le maniement efficace et intelligent du texte en langue naturelle dans les phases critiques du traitement du document, comme l'identification, le résumé, l'indexation, l'extraction et la présentation des faits, le stockage et la recherche des documents ainsi que la traduction, et permettra également plus de puissance et de commodité dans la communication en langue naturelle avec les machines.

Basé à Cambridge (Royaume-Uni) et Grenoble (France), le Xerox Research Centre Europe (XRCE) est lui aussi un laboratoire de recherche de la société XEROX. Les travaux menés ont pour but d'améliorer la productivité sur le lieu de travail, grâce à la mise en oeuvre de nouvelles technologies centrées sur le document.

Un des projets du XRCE est Théories et technologies multilingues (MLTT), qui étudie l'analyse et la génération de textes pour une grande variété de langues (allemand, anglais, arabe, espagnol, français, italien, russe, etc.). L'équipe du MLTT crée des outils de base pour l'analyse linguistique multilingue, tels que analyseurs morphologiques, étiqueteurs morpho-syntaxiques, plate-formes pour le parsing et la génération, ou encore outils d'analyse de corpus. Ces outils sont utilisés pour décrire diverses langues ainsi que les relations d'une langue à l'autre. Les projets en cours incluent des analyses syntaxiques à états finis pour le français et l'allemand, une grammaire LFG (lexical functional grammar - grammaire fonctionnelle lexicale) du français ainsi que des projets en recherche documentaire multilingue, en génération et en traduction.

Créée en 1979, l'American Association for Artificial Intelligence (AAAI) (Association américaine pour l'intelligence artificielle) est une société scientifique à but non lucratif visant une meilleure compréhension scientifique des mécanismes sous-jacents de la pensée et du comportement intelligent et de leur incorporation dans les machines. L'AAAI a également pour but de favoriser la compréhension de l'intelligence artificielle par le grand public, d'améliorer l'enseignement et la formation des praticiens en intelligence artificielle, et de procurer des conseils aux planificateurs de recherches et financeurs sur l'importance des développements en cours sur l'intelligence artificielle, les possibilités qui en découlent et les orientations futures.

Rattaché à l'Université de Genève (Suisse), l'Institut Dalle Molle pour les études sémantiques et cognitives (ISSCO) mène des recherches théoriques et appliquées en linguistique computationnelle et en intelligence artificielle.

Créé par la Fondation Dalle Molle en 1972 pour mener des recherches en cognition et en sémantique, l'Institut en est venu à se spécialiser dans le traitement de la langue naturelle et, en particulier, dans le traitement multilingue des langues dans un certain nombre de domaines: traduction automatique,

environnement linguistique, génération multilingue, traitement du discours, collection de données, etc. Si l'université de Genève procure un soutien administratif et une infrastructure à l'ISSCO, la recherche est uniquement financée par des subventions et des contrats avec des organismes publics et privés.

L'Institut est multidisciplinaire et multinational, avec un petit groupe de permanents complété par un certain nombre de personnes sous contrat (spécialisées en informatique, linguistique, mathématiques, psychologie et philosophie) restant de six mois à deux ans, ce qui permet une grande flexibilité et un échange continu d'idées.

L'International Committee on Computational Linguistics (ICCL) (Comité international de linguistique computationnelle) organise l'International Conference on Computational Linguistics (COLING) (Conférence internationale de linguistique computationnelle). L'ICCL a été mis sur pied dans les années 60 par David Hays en tant qu'organisme permanent

"pour organiser des conférences internationales en linguistique computationnelle d'une manière originale, sans secrétariat permanent, inscriptions ou financement propre. Pour ces raisons et d'autres, cette démarche était en avance sur son temps. Les COLINGs ont toujours été des lieux de rendez-vous caractérisés par une atmosphère agréable, plutôt que des rassemblements à l'efficacité clinique se déroulant dans un hôtel d'aéroport. [...] Ces dernières années, l'ACL [Association for Computational Linguistics - Association de linguistique computationnelle] a beaucoup contribué à la réalisation des compte-rendus de conférences et à leur diffusion."

### 5.3. Ingénierie du langage

Lancé en janvier 1999 par la Commission européenne, le site HLTCentral (HLT: Human Languages Technologies - Technologies des langues humaines) propose une courte définition de l'ingénierie du langage:

"L'ingénierie du langage permet de vivre en toute convivialité avec la technologie. Nous pouvons utiliser notre connaissance du langage pour développer des systèmes capables de reconnaître à la fois la parole et l'écrit, de comprendre un texte suffisamment en profondeur pour être capable de sélectionner des informations, de le traduire dans différentes langues et de générer aussi bien un discours oral qu'un texte imprimé.

L'application de ces technologies nous permet de repousser les limites actuelles de notre utilisation du langage. Les systèmes à commande vocale sont appelés à jouer un rôle prépondérant et à faire partie intégrante de notre vie quotidienne."

Une présentation très complète de l'ingénierie du langage est également proposée dans L'ingénierie linguistique, ou comment exploiter la puissance du langage.

Créé par le projet LINGLINK, HLTCentral (HLT: Human Language Technologies - Technologies du langage humain) veut rassembler les ressources en technologies des langues présentes sur le Web: informations, actualités, fichiers à télécharger, liens, événements, groupes de discussion et études commissionnées (commerce électronique, télécommunications, localisation, etc.).

Le secteur HLT fait partie du programme IST (Information Society Technologies - Technologies de la société de l'information) lancé par la Commission européenne pour la période 1999-2002. Il succède à Ingénierie linguistique, développé entre 1992 et 1998 par le Programme d'applications télématiques. Son but était de faciliter l'utilisation d'applications télématiques et d'augmenter les possibilités de communication entre langues européennes. Les travaux de RTD (recherche et développement technologique) concernaient principalement des projets pilotes intégrant les technologies de la langue dans les applications et services d'information et de communication.

FRANCIL (Réseau francophone de l'ingénierie de la langue) est un programme mis en place en juin 1994 par l'Agence universitaire de la francophonie (AUPELF-UREF) pour renforcer ses activités dans le domaine du génie linguistique, défini ainsi sur le site:

"Le Génie Linguistique est une partie du traitement électronique de l'information. Dans ce contexte, le traitement automatique des langues est un secteur en plein développement. Il comprend les recherches et développement en matière d'analyse et de génération de textes, de reconnaissance, de compréhension et de synthèse de la parole. Il inclut les applications relatives à la gestion de documents, à la communication entre l'humain et la machine, à l'aide à la rédaction, à la traduction assistée par ordinateur. Il comporte des enjeux de type industriel et économique, de type scientifique et technologique mais présente aussi une dimension culturelle très spécifique. Il est toujours préférable de maîtriser la langue qui sert à exprimer les résultats des recherches, surtout si cette recherche porte

sur la langue elle-même. La langue française et la francophonie sont donc très directement concernées par cet enjeu."

Comme son nom l'indique, la Multilingual Application Interface for Telematic Services (MAITS) (Interface pour les applications multilingues des services télématiques) est un consortium formé pour développer un API (applications programming interface - interface pour la programmation des applications) pour les applications multilingues des services télématiques.

#### 5.4. Internationalisation et localisation

"Vers la communication sur Internet dans toutes les langues..." Babel est un projet conjoint d'Alis Technologies et de l'Internet Society traitant de l'internationalisation d'Internet. Son site multilingue (allemand, anglais, espagnol, français, italien, portugais et suédois) comprend deux grands secteurs: langues (les langues du monde; glossaire typographique et linguistique; Francophonie), et Internet et multilinguisme (développer votre site web multilingue; le codage des écritures du monde).

La Localisation Industry Standards Association (LISA) (Association pour les normes de l'industrie de la localisation) est une organisation majeure pour l'industrie de localisation et d'internationalisation. Ses 130 membres comprennent des éditeurs de logiciels, des fabricants de matériel, des vendeurs de services de localisation, et un nombre croissant de sociétés venant des secteurs voisins de technologie de l'information. La mission de LISA est de promouvoir l'industrie de la localisation et de l'internationalisation et de procurer un mécanisme et des services permettant aux sociétés d'échanger et de partager l'information dans le développement de processus, outils, technologies et modèles de sociétés en rapport avec la localisation, l'internationalisation et les domaines voisins. Son site est hébergé par l'Université de Genève, en Suisse.

W3C Internationalization/Localization est un secteur du World Wide Web Consortium (W3C), consortium international de l'industrie fondé en 1994 pour développer les protocoles communs du World Wide Web. Le site donne en particulier une définition des protocoles utilisés pour l'internationalisation et la localisation: HTML (hypertext markup language), jeu 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 (American standard code for information interchange). Le site propose aussi une aide pour créer un site multilingue.

## 6. SELECTION DE SITES

[Dictionnaires - exemples / Dictionnaires - répertoires / Ingénierie de la langue / Internationalisation et localisation / Langue française - promotion et recherche / Langues - enseignement / Langues - répertoires / Langues appliquées / Linguistique computationnelle / Multilinguisme - chiffres / Multilinguisme en Europe / Outils linguistiques - répertoires / Terminologie / Terminologie - informatique / Terminologie - Internet / Traduction / Traduction automatique - associations / Traduction automatique - recherche / Traduction automatique - services gratuits sur le Web / Traduction automatique - sociétés]

#Dictionnaires - exemples

= Dictionnaire universel francophone en ligne

Publié par Hachette et l'AUPELF-UREF (Agence francophone pour l'enseignement supérieur et la recherche). "Voici enfin présentés, 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." [français]

= The Internet Dictionary Project (IDP)

Créé en 1995 par Tyler Chambers, un projet coopératif en cours pour la constitution de dictionnaires en accès libre sur le Web - de l'anglais vers d'autres langues (allemand, espagnol, français, italien, latin et portugais). Tout internaute peut y participer. A suivre pour juger de la qualité. [menu en anglais]

= The Logos Dictionary

Un dictionnaire multilingue d'environ 7,5 millions d'entrées géré par Logos, une société de traduction internationale dont le siège est à Modène (Italie) et qui met en accès libre sur son site tous les outils utilisés par ses traducteurs. Outre le Logos Dictionary, on a aussi la Wordtheque, une base de données multilingue de 400 millions de mots présentés dans leur contexte (le texte de romans, documents techniques et écrits divers), Linguistic Resources (Ressources linguistiques), qui regroupe 650



glossaires, et enfin le Universal Conjugator (Conjugaison universelle), qui propose des tableaux de conjugaison dans 17 langues différentes. [anglais]

= Merriam-Webster Online: The Language Center

Un grand éditeur américain de dictionnaires met en accès libre deux ouvrages de référence, le Webster Dictionary et le Webster Thesaurus. [anglais]

#Dictionnaires - répertoires

= Dictionnaires électroniques

Préparé par la Section française des Services linguistiques centraux (SLC-f) de la Chancellerie fédérale suisse. Propose cinq rubriques: dictionnaires monolingues, dictionnaires bilingues, dictionnaires multilingues, abréviations et acronymes, et informations géographiques. [français]

= OneLook Dictionaries

Créé en avril 1996 par Robert Ware, un moteur de recherche rapide puisant dans 2,5 millions de mots disponibles dans 530 dictionnaires traitant des domaines suivants: affaires, informatique/Internet, médecine, religion, sciences, sports, technologie, généralités et argot. [anglais]

= Travlang's Translating Dictionaries

Sur le site de Michael C. Martin, dédié à la fois aux voyages et aux langues, une section donnant accès à des dictionnaires gratuits dans diverses langues (afrikaans, allemand, danois, espagnol, espéranto, finnois, français, frison, hollandais, hongrois, italien, latin, norvégien, portugais et tchèque). [anglais]

= A Web of Online Dictionaries

Oeuvre de Robert Beard, un index des dictionnaires en ligne dans plus de 170 langues différentes, auquel s'ajoutent d'autres sections: dictionnaires multilingues, dictionnaires anglophones spécialisés, thésaurus et vocabulaires, grammaires en ligne, et outils linguistiques pour non spécialistes. [anglais]

#Ingénierie de la langue

= HLTCentral (HLT: Human Language Technologies)

Lancé en janvier 1999 par le projet LINGLINK de la Commission européenne, HLTCentral veut rassembler les ressources en technologies des langues présentes sur le Web: informations, actualités, fichiers à télécharger, liens, événements, groupes de discussion et études commissionnées (commerce électronique, télécommunications, localisation, etc.). HLT fait partie du programme IST (Information Society Technologies) lancé par la Commission européenne pour la période 1999-2002, et qui succède au programme Ingénierie linguistique (1992-1998). [anglais]

= Multilingual Application Interface for Telematic Services (MAITS)

Consortium formé pour développer un API (applications programming interface = interface de programmation d'application) pour les applications multilingues des services télématiques. [français, résumés dans plusieurs langues]

= **FRANCIL**

FRANCIL (Réseau francophone de l'ingénierie de la langue) est un programme mis en place en juin 1994 par l'Agence universitaire de la francophonie (AUPELF-UREF) pour renforcer ses activités dans le domaine du génie linguistique, notamment le traitement automatique des langues. [français]

#Internationalisation et localisation

= Babel

Projet conjoint d'Alis Technologies et de l'Internet Society traitant de l'internationalisation d'Internet. Propose deux grands secteurs: 1) Langues (Les langues du monde - Glossaire typographique et linguistique - La Francophonie), 2) Internet et multilinguisme (Développer votre site web multilingue - Le codage des écritures du monde). [allemand, anglais, espagnol, français, italien, portugais et suédois]

= Localisation Industry Standards Association (LISA)

Le site d'une organisation qui oeuvre pour l'industrie de localisation et d'internationalisation. Ses 130

membres comprennent des éditeurs de logiciels, des fabricants de matériel, des vendeurs de services de localisation, et un nombre croissant de sociétés venant des secteurs voisins des technologies de l'information. LISA procure un mécanisme et des services permettant aux sociétés d'échanger et partager l'information dans le développement de processus, outils, technologies et modèles de sociétés. [anglais]

= W3C Internationalization / Localization

Un secteur du World Wide Web Consortium (W3C), consortium industriel international fondé en 1994 pour développer les protocoles communs du World Wide Web. Le site donne notamment une définition des protocoles utilisés pour l'internationalisation et la localisation: HTML (hypertext markup language), HTTP (hypertext transfer protocol), URL (uniform resource locator) et autres identificateurs incluant des caractères non-ASCII (American standard code for information interchange), etc. Donne aussi des conseils pour créer un site multilingue. [anglais]

#Langue française - promotion et recherche

= Agence de la francophonie

Créée en 1970 pour regrouper 21 états francophones, l'Agence de la Francophonie en compte aujourd'hui 47. "Instrument de coopération multilatérale née d'un idéal, celui de créer une communauté qui fasse entendre sa voix dans le concert des nations, elle participe aujourd'hui à l'avènement d'un Secrétariat général de la Francophonie." [français]

= Agence universitaire de la Francophonie

Opérateur direct du Sommet des Chefs d'Etat et de gouvernement des pays ayant le français en partage, cet organisme a le mandat officiel de contribuer à la construction francophone en consolidant un espace scientifique de langue française animé par ses principaux acteurs, les établissements, les enseignants, les chercheurs et les étudiants. [français]

##Belgique

= Maison de la Francité

Association belge à but non lucratif subventionnée par la Commission communautaire française, la Maison de la Francité présente la réalité socio-linguistique de Bruxelles, seconde capitale internationale de langue française après Paris, et elle agit pour la défense et la promotion de la langue française à Bruxelles et au sein de la Communauté française Wallonie-Bruxelles. [français]

##France

= Délégation générale à la langue française (DGLF)

La DGLF a pour mission de veiller à l'emploi et à la promotion du français sur le territoire national, favoriser son utilisation comme langue de communication internationale et développer le plurilinguisme, garant de la diversité culturelle. [français]

= Institut national de la langue française (INaLF)

L'INaLF est une branche du CNRS (Centre national de la recherche scientifique). Ses recherches portent sur la langue française sous tous ses aspects, notamment le discours littéraire du 14e au 20e siècle (contenu, sémantique et thématique), la langue courante (langue écrite, langue parlée et argot), le discours scientifique et technique et ses ressources terminologiques. [français]

##Québec

= Centre d'expertise et de veille inforoutes et langues (CEVEIL)

Créé en 1995, un organisme québécois à but non lucratif dont la mission est de mieux cerner la problématique de l'utilisation et du traitement des langues sur les inforoutes - dans une optique plus spécifiquement francophone - via l'institution d'activités de veille stratégique et la création d'un réseau d'échanges et d'expertise. [français]

= Office de la langue française (OLF)

Organisme gouvernemental québécois dont le mandat est d'une part de veiller à l'implantation et au maintien du français dans le monde du travail et des affaires et dans les services administratifs, d'autre part de définir et conduire la politique québécoise en matière de linguistique et de terminologie. [français, anglais]

## #Langues - enseignement

= CTI Centre for Modern Languages

Inclus dans l'Institut des langues de l'Université d'Hull (Royaume-Uni), le CTI Centre for Modern Languages (CTI: Computer in Teaching Initiative) vise à promouvoir l'utilisation des ordinateurs dans l'apprentissage et l'enseignement des langues. Le Centre procure des informations sur la manière dont l'apprentissage des langues assisté par ordinateur peut être intégré à des cours existants, et il offre un soutien aux professeurs utilisant l'informatique dans l'enseignement qu'ils dispensent. Voir aussi la très intéressante rubrique: Internet Resources for Language Teachers and Learners. [anglais]

= **EUROCALL**

EUROCALL (European Association for Computer-Assisted Language Learning = Association européenne pour l'apprentissage des langues assisté par ordinateur) regroupe des professionnels de l'enseignement des langues exerçant en Europe et dans le monde entier. Le prochain congrès annuel, EUROCALL'99, aura lieu du 15 au 18 septembre 1999 à l'Université de Franche-Comté (Besançon, France). Comme chaque année, il fera le point sur les recherches et les applications informatiques dans le domaine de l'apprentissage des langues. [anglais]

= WELL (Web Enhanced Language Learning)

Un projet anglais en cours (1997-2000), qui est en train de mettre sur pied: a) un vaste programme de séminaires, b) une liste d'études de cas sur l'apprentissage et l'enseignement des langues par le biais du Web, c) un répertoire complet de ressources web de qualité dans douze langues différentes. [anglais]

## #Langues - répertoires

= Ethnologue: Languages of the World

Edité par Barbara F. Grimes, la version web d'un catalogue très documenté répertoriant 6.700 langues, avec de multiples critères de recherche. Cet ouvrage de référence (13e édition) est également disponible sur papier et sur CD-ROM. [anglais]

= European Minority Languages

Gérée par Caoimhín P. Ó Donnáil sur le site de l'Institut Sabhal Mór Ostaig (Ile de Skye, Ecosse), une liste de langues minoritaires - ou rendues minoritaires - disponible par ordre alphabétique et par famille linguistique. Ce site procure aussi des liens avec d'autres répertoires dans le monde entier (Other Directories Worldwide). [anglais, gaélique]

= The Languages of the World by Computers and the Internet

Créé en décembre 1995 par Yoshi Mikami, de Asia Info Network, ce site est connu aussi sous le nom de Logos Home Page ou Kotoba Home Page. Il donne, pour chaque langue, un bref historique, les caractéristiques, le système d'écriture, le jeu de caractères et la configuration du clavier pour l'utilisation de l'ordinateur et d'Internet. [anglais, japonais]

## #Langues appliquées

= Language today

Un magazine passionnant pour ceux qui travaillent dans les langues appliquées: traducteurs, interprètes, terminologues, lexicographes et rédacteurs techniques. Ce magazine est une réalisation commune de Logos, qui procure le site web, et Praetorius, société de traduction et service d'expertise britannique. [anglais]

## #Linguistique computationnelle

= The Association for Computational Linguistics (ACL)

Un organisme international à la fois scientifique et professionnel rassemblant ceux qui travaillent sur les problèmes de la langue naturelle et de la computation. Publiée par la MIT Press, la revue trimestrielle de l'ACL, Computational Linguistics (ISSN 0891-2017) est un forum de premier plan dans le domaine de la linguistique computationnelle et du traitement de la langue naturelle. La branche européenne de l'ACL est l'European Chapter of the Association for Computational Linguistics (EACL). [anglais]

= Institut Dalle Molle pour les études sémantiques et cognitives (ISSCO)

Rattaché à l'Université de Genève (Suisse), l'ISSCO mène des recherches théoriques et appliquées en linguistique computationnelle et en intelligence artificielle. Il est spécialisé dans le traitement de la langue naturelle et, en particulier, dans le traitement multilingue des langues en traduction automatique, environnement linguistique, génération multilingue, traitement du discours, collecte des données, etc. [anglais]

= International Committee on Computational Linguistics (ICCL) / COLING

Créé pour organiser des conférences internationales en linguistique computationnelle - dénommées COLING (Computational Linguistics) - "d'une manière originale, sans secrétariat permanent, inscriptions ou financement propre. Pour ces raisons et d'autres, cette démarche était en avance sur son temps. Les COLINGs ont toujours été des lieux de rendez-vous caractérisés par une atmosphère agréable, plutôt que des rassemblements à l'efficacité clinique se déroulant dans un hôtel d'aéroport. "Les compte-rendus de conférences sont réalisés avec l'aide de l'Association for Computational Linguistics. [allemand, anglais, français, italien, polonais, portugais, romanche]

= Le Laboratoire de recherche appliquée en linguistique informatique (RALI)

Basé à Montréal (Québec), le RALI réunit des informaticiens et des linguistes qui étudient le traitement automatique de la langue tant par des méthodes symboliques "classiques" que par de nouvelles méthodes probabilistes. Ses domaines de compétence sont les suivants: outils d'aide à la traduction, appariement automatique de textes, génération automatique de texte, réaccentuation automatique, recherche d'information aidée par des outils linguistiques, extraction d'information, identification de la langue et du codage, transducteurs à états finis, et corpus de texte enrichis. [français, anglais]

= Natural Language Group (NLG) at USC/ISI

Un département de l'Information Sciences Institute (ISI) de l'University of Southern California (USC). Traite de plusieurs aspects du traitement de la langue naturelle: traduction automatique, résumé automatique de texte, accès multilingue aux verbes et gestion du texte, développement de taxonomies de concepts (ontologies), discours et génération de texte, élaboration de grands lexiques pour plusieurs langues, et communication multimédia. [anglais]

= Text Encoding Initiative (TEI)

Un projet international pour mettre au point des directives sur l'encodage des textes électroniques à destination de la recherche. Un consortium a été formé récemment par les trois sponsors de ce projet (Association for Computers and the Humanities, Association for Literary and Linguistic Computing, Association for Computational Linguistics). [anglais]

= University of Essex - CL/MT Research Group

Le CL/MT Research Group (CL/MT: Computational Linguistics / Machine Translation) est un groupe de recherche du Département des langues et de linguistique de l'Université d'Essex (Royaume-Uni). [anglais]

= Xerox Palo Alto Research Center (PARC)

Le centre de recherche de la société Xerox dans la Silicon Valley (Californie) mène au moins deux projets relatifs aux langues: Inter-Language Unification (ILU) et Natural Language Theory and Technology (NLTT). [anglais]

= Xerox Research Centre Europe (XRCE)

Le centre de recherche européen de la société Xerox, basé à Cambridge (Royaume-Uni) et Grenoble (France). Les travaux menés ont pour but d'améliorer la productivité sur le lieu de travail, grâce à la mise en oeuvre de nouvelles technologies centrées sur le document, notamment dans le programme Théories et technologies multilingues (MLTT). [allemand, anglais, français, italien]

[Voir aussi la section: Traduction automatique - recherche]

#Multilinguisme - chiffres

= Global Reach

Sur le site de Global Reach, qui traduit les sites web commerciaux tout en les adaptant aux besoins de chaque pays, des statistiques détaillées sur la répartition des langues sur le Web. [anglais]

= Palmarès des langues de la Toile

Menée par Babel, un projet conjoint d'Alis Technologies et de l'Internet Society, la première étude d'ensemble sur la répartition réelle des langues sur Internet. Comme elle date de juin 1997, on attend impatiemment une nouvelle version... [français, anglais]

#Le multilinguisme en Europe

= Association européenne pour les ressources linguistiques (ELRA)

Etablie à Luxembourg en février 1995, une association à but non lucratif qui a pour but d'une part de fournir une organisation centralisée pour la validation, la gestion et la distribution des ressources et outils en parole, texte et terminologie, d'autre part de promouvoir leur utilisation au sein de la section européenne de recherche et développement en télématique. [français, anglais]

= European Network in Language and Speech (ELSNET)

Regroupe plus d'une centaine d'institutions universitaires et industrielles. L'objectif technologique commun aux participants de l'ELSNET est de construire des systèmes multilingues pour la parole et le langage naturel, avec couverture illimitée de la langue parlée et écrite. [anglais]

= Language Futures Europe

Paul Treanor propose des répertoires sur la politique linguistique, le multilinguisme, les structures linguistiques globales, la domination de l'anglais, la politique européenne, les politiques nationales, les sites de recherche et le "mouvement monolingue" émergeant aux Etats-Unis. [anglais]

= Multilingual Information Society (MLIS)

Un programme de la Commission européenne lancé en novembre 1996 à destination des entreprises, des organismes du secteur public, des industries des langues et des citoyens. Ses objectifs: soutenir la réalisation d'une infrastructure favorisant des ressources linguistiques européennes, stimuler le développement des industries des langues, et favoriser l'utilisation d'outils linguistiques dans le secteur public européen. [français, disponible dans les onze langues de la Communauté européenne]

#Outils linguistiques - répertoires

= The Human-Languages Page (H-LP)

Créé par Tyler Chambers en mai 1994, un catalogue détaillé de 1.800 ressources linguistiques dans plus de 100 langues différentes, réparties en différentes sections: langues et littérature, écoles et institutions, ressources linguistiques, produits et services, organismes, emplois et stages, dictionnaires et cours de langues. [anglais]

= The LINGUIST List

Proposé par l'Eastern Michigan University et la Wayne State University (USA), une série de répertoires sur les ressources linguistiques: profession (conférences, associations linguistiques, programmes, etc.), recherche et soutien à la recherche (articles, résumés de mémoires, projets, bibliographies, dossiers, textes), publications, pédagogie, ressources linguistiques (langues, familles linguistiques, dictionnaires, information régionale) et soutien informatique (polices de caractères et logiciels). Propose aussi un très bon centre de documentation virtuel (Virtual Library), un moteur de recherche axé sur les langues (Search for LINGUIST Issues) et un moteur de recherche plus général pour l'ensemble du site (Search the LINGUIST List web site). [anglais]

= Multilingual Tools and Services

Proposé par la Commission européenne dans le cadre du programme Télématique pour les bibliothèques, répertorie une série de dictionnaires, instruments d'aide multilingues, projets, moteurs de recherche par langue, banques de données terminologiques, thésaurus et systèmes de traduction. [anglais]

= Speech on the Web

Géré par l'Institut des sciences phonétiques d'Amsterdam (Pays-Bas), un répertoire de sites organisés en différentes sections: congrès, réunions et ateliers, liens et listes, phonétique et parole, traitement de la langue naturelle, sciences cognitives et intelligence artificielle, linguistique computationnelle, dictionnaires, lettres d'information électroniques, revues et publications. [anglais]

= Travlang

Créé par Michael C. Martin en 1994, un excellent site dédié à la fois aux voyages et aux langues. Foreign Languages for Travelers donne la possibilité d'apprendre 60 langues différentes sur le Web. Translating Dictionaries donne accès à des dictionnaires gratuits dans diverses langues (afrikaans, allemand, danois, espagnol, espéranto, finnois, français, frison, hollandais, hongrois, italien, latin, norvégien, portugais et tchèque). Le site offre aussi de nombreux liens vers des services de traduction, des écoles de langue, des librairies multilingues, etc. [anglais]

#Terminologie

= Eurodicautom

Gérée par le Service de traduction de la Commission européenne, une base de données terminologique disponible dans les onze langues officielles de l'Union européenne (allemand, anglais, danois, espagnol, finnois, grec, hollandais, italien, latin, portugais et suédois). Eurodicautom était à l'origine un outil de travail destiné aux traducteurs de la Commission. Disponible sur le Web, il est maintenant consulté par des professionnels des langues dans le monde entier. [menu en anglais]

= **ILOTERM**

Gérée par l'Unité de terminologie et de références du Service des documents officiels (OFFDOC) du Bureau international du Travail (BIT), une base de données terminologique quadrilingue (allemand, anglais, espagnol et français) relative au travail et aux questions sociales. Inclut les noms de nombreux programmes et organismes. [menu en anglais, espagnol et français]

= ITU Telecommunication Terminology Database (TERMITE)

Géré par la Section de traduction de l'Union internationale des télécommunications (UIT), une base terminologique quadrilingue (environ 60.000 entrées en anglais, espagnol, français et russe) contenant tous les termes inclus dans les glossaires de l'UIT imprimés depuis 1980, ainsi que des termes plus récents en rapport avec les différentes activités de l'Union. [menu en anglais]

= The WHO Terminology Information System (WHOTERM)

Géré par l'Organisation mondiale de la santé (OMS), un système d'information terminologique trilingue (anglais, espagnol et français) constitué à partir des expressions et concepts des documents de l'OMS et permettant aussi d'enrichir les terminologies nouvelles, promouvoir leur normalisation et garantir leur diffusion. [menu en anglais]

#Terminologie - informatique

= FOLDOC - Free On-Line Dictionary of Computing

Par Denis Howe, 12.574 termes listés le 3 juillet 1999, plus une série de liens. [anglais]

= Glossaire informatique des termes de la Commission ministérielle de terminologie

"Ce document est le résultat d'une compilation des divers arrêtés issus des travaux de la Commission ministérielle de terminologie informatique ainsi que du projet d'arrêté qui était en cours lorsque le dispositif terminologique a fait l'objet d'une profonde réforme. Une compilation initiale avait été réalisée par Daniel Duthil de l'APP, la version actuelle et la présentation ont été commises par Philippe Deschamp de l'INRIA. Tous deux étaient membres de la dite commission." [français]

= Le signet: la référence branchée en technologies de l'information

6 mille fiches bilingues (français-anglais) proposées par l'Office de la langue française (Québec). [français]

#Terminologie - Internet

= L'ABC des réseaux

Conçu et réalisé par l'association Passerelles - Langue française et nouvelles technologies. [français]

= Glossaire de termes relatifs à Internet

Un glossaire officieux de Jean-Karim Benzineb, traducteur au Conseil de l'Europe. [français]

= Lexique des néologismes Internet

Par Guy Brand et Jean-Pierre Kuypers. Rassemble quelques équivalences du vocabulaire technique lié aux réseaux informatiques et aux services qu'ils supportent. Ces termes sont communément utilisées par des traducteurs et adaptateurs francophones de logiciels diffusés sur Internet. [français]

= NetGlos (The Multilingual Glossary of Internet Terminology)

Créé en 1995 à l'initiative du WorldWide Language Institute, NetGlos est le projet commun d'un certain nombre de traducteurs et autres professionnels des langues. Ce glossaire est préparé dans les langues suivantes: allemand, anglais, chinois, croate, espagnol, français, grec, hébreu, hollandais/flamand, italien, maori, norvégien et portugais. [menu en anglais]

= Inventaire des terminologies dans Internet

Proposé par l'Office de la langue française (Québec). [français]

#Traduction

= Aquarius

Créé par Vorontsoff, Wesseling & Partners (Amsterdam, Pays-Bas), un répertoire non commercial de traducteurs et interprètes - 7.861 traducteurs et 1.164 sociétés de traduction au 4 juillet 1999. Permet de localiser et de contacter les personnes et sociétés directement, sans intermédiaire ni agence. La recherche est possible par lieu, combinaison de langues et spécialité. [anglais]

= Global Reach (Euro-Marketing Associates)

Fondé par Bill Dunlap, Global Reach est un service commercial à destination des sociétés américaines et européennes voulant étendre leur présence sur Internet à l'international. Les prestations incluent la traduction et l'adaptation du site web dans d'autres langues, la promotion de ce site et l'utilisation de bandeaux publicitaires nationaux pour augmenter la consultation locale. [allemand, anglais, danois, espagnol, français, italien, néerlandais, portugais et suédois]

#Traduction automatique - associations

= Association for Machine Translation in the Americas (AMTA)

Association à la disposition de tous ceux qui s'intéressent à la traduction automatique en Amérique latine, au Canada et aux Etats-Unis. Ses membres comprennent des demandeurs de traductions, des développeurs de systèmes commerciaux, des chercheurs, des sponsors et des personnes impliquées dans la recherche en traduction automatique. [anglais]

= European Association for Machine Translation (EAMT)

Basée à Genève (Suisse), une organisation à la disposition de ceux qui s'intéressent à la traduction automatique et aux outils de traduction, y compris les utilisateurs de ces techniques, les développeurs et les chercheurs. [anglais]

= Asia-Pacific Association for Machine Translation (AAMT)

Créée en 1991 et d'abord appelée Japan Association for Machine Translation, comprend des chercheurs, des fabricants et des utilisateurs de systèmes de traduction automatique. Participe au développement des technologies de traduction automatique, ainsi qu'à leur amélioration, leur enseignement et leur diffusion. [anglais]

#Traduction automatique - recherche

= Center for Machine Translation (CMT)

Fondé en 1986, un centre de recherche inclus dans le nouvel Institut des technologies des langues de la Carnegie Mellon University (CMU, Pittsburgh, Pennsylvanie, USA). Il est spécialisé dans les technologies de traitement de la langue naturelle, particulièrement la traduction automatique multilingue de haute qualité. [anglais]

= Computing Research Laboratory (CRL)

Ce centre de recherche de la New Mexico State University (NMSU) se consacre à la recherche de base et au développement de logiciels dans les applications informatiques avancées du traitement de la langue naturelle, de l'intelligence artificielle et de la conception d'interfaces graphiques pour les utilisateurs. Les applications développées à partir de cette recherche de base incluent un ensemble de configurations de traduction automatique, extraction d'information, acquisition du savoir,

enseignement intelligent, et systèmes de traduction pour poste de travail. [anglais]

= Groupe d'étude pour la traduction automatique (GETA)

Au sein du Laboratoire CLIPS (Communication langagière et interaction personne-système) de la Fédération IMAG (France), une équipe pluridisciplinaire formée d'informaticiens et de linguistes. Les thèmes de recherche du GETA concernent tous les aspects théoriques, méthodologiques et pratiques de la traduction assistée par ordinateur (TAO du réviseur, du traducteur et du rédacteur), et plus généralement de l'informatique multilingue. [français]

= Machine Translation at the Pan American Health Organization: SPANAM and ENGSPAN

Un système de traduction automatique développé par les linguistes computationnels, les traducteurs et le programmeur systèmes de l'Organisation panaméricaine de la santé (OPS) (qui est le Bureau régional de l'Organisation mondiale de la santé pour les Amériques, situé à Washington, D.C., USA). Depuis 1980, le service de traduction utilise SPANAM (de l'espagnol vers l'anglais) et ENGSPAN (de l'anglais vers l'espagnol), ce qui lui a permis de traiter plus de 25 millions de mots dans les deux langues de travail de l'OPS. [anglais]

= TTT.org -Translation, Theory and Technology

Géré par le Département de linguistique du Groupe de recherche sur la traduction de l'Université Brigham Young (Utah, USA), TTT.org procure des informations sur la théorie et la technologie de la langue, particulièrement dans le domaine de la traduction. La technologie de la traduction inclut les outils de traduction pour poste de travail et la traduction automatique. TTT.org s'intéresse aussi aux normes d'échanges de données permettant à divers outils de fonctionner en lien les uns avec les autres, ce qui permet l'intégration d'outils de la part de vendeurs multiples dans la chaîne multilingue de production de documents. [anglais]

= Universal Networking Language (UNL)

Sous l'égide de l'Université des Nations Unies (UNU, Tokyo), un projet de "métalangage numérique" pour l'encodage, le stockage, la recherche et la communication d'information multilingue indépendamment d'une langue-source (et donc d'un système de pensée) donnée. 120 chercheurs de par le monde travaillent sur un projet multilingue comportant 17 langues (allemand, anglais, arabe, brésilien, chinois, espagnol, français, hindou, indonésien, italien, japonais, letton, mongolien, russe, swahili et thaï). [site en anglais et japonais - résumé du projet en allemand et en anglais]

#Traduction automatique - services gratuits sur le Web

= Altavista Translation / Babel Fish

Proposé depuis décembre 1997 par le moteur de recherche AltaVista, un service de traduction automatisée de l'anglais vers cinq autres langues (allemand, espagnol, français, italien et portugais), et vice versa. Alimenté par des dictionnaires multilingues contenant plus de 2,5 millions de termes, ce service, gratuit, instantané et approximatif, a été mis en place par Systran, société pionnière dans le domaine de la traduction automatique. [anglais]

= FreeTranslation.com

Proposé par Transparent Language, un service de traduction automatique rapide et approximative d'une page web ou d'un texte personnel. Langues proposées: 1) l'anglais vers l'allemand, l'espagnol, le français, l'italien et le portugais, 2) l'allemand, l'espagnol et l'italien vers l'anglais. [anglais]

= Go Translator

Un logiciel de la société Systran en accès libre sur le Web. Permet de traduire une page web ou un texte personnel, avec les mêmes combinaisons de langues qu'Altavista Translation (anglais vers allemand, espagnol, français, italien et portugais, et vice versa). [anglais]

#Traduction automatique - sociétés

= Alis Technologies

Basée à Montréal (Québec), Alis Technologies développe et commercialise des solutions et services de traitement linguistique au moyen de logiciels de traduction qui transforment des systèmes informatiques unilingues en outils multilingues. [site en anglais et français - pages en espagnol et japonais pour la traduction automatique]



= Lernout & Hauspie (L&H)

Basé à Ypres (Belgique) et Burlington (Massachusetts, USA), Lernout & Hauspie (L&H) est une grosse société de produits, technologies et services incluant la reconnaissance automatique de la langue, la compression numérique de la parole, le passage du texte à la parole, la compression et la traduction. [anglais]

= Logos Corporation

Avec des bureaux sur les deux côtes des Etats-Unis et en Europe, Logos Corporation fournit des solutions de traduction et de localisation à des sociétés représentant une large gamme d'industries, y compris les technologies de l'information et les télécommunications. [allemand, anglais, chinois simplifié, chinois traditionnel, coréen, espagnol, italien, japonais et portugais]

= Softissimo

Edite et diffuse des logiciels de traduction automatique, d'apprentissage des langues et de dictionnaires (Eurodico, Grand Collins bilingue et Collins English Dictionary). [allemand, anglais, français]

= **SYSTRAN**

SYSTRAN (acronyme de: System Translation) est une société spécialisée dans les logiciels de traduction automatique. Son siège est situé à Soisy-sous-Montmorency (France). C'est sa succursale, située à La Jolla (Californie), qui assure les ventes, le marketing et le développement des logiciels. La réalisation la plus connue de la société est AltaVista Translation / Babel Fish, mis en place en décembre 1997 sur le site du moteur de recherche AltaVista, et qui permet de traduire les pages web de l'anglais vers les langues suivantes: allemand, français, espagnol, italien et portugais, et vice versa. [anglais]

## **7. INDEX DES SITES ET PAGES WEB**

Agence de la francophonie

Alis Technologies

AltaVista Translation

American Association for Artificial Intelligence (AAAI)

Aquarius

ARTFL Project (ARTFL: American and French Research on the Treasury of the French Language)

Asia-Pacific Association for Machine Translation (AAMT)

Association européenne pour les ressources linguistiques (ELRA)

Association for Computational Linguistics (ACL)

Association for Machine Translation in the Americas (AMTA)

Babel / Alis Technologies & Internet Society

CAPITAL (Computer-Assisted Pronunciation Investigation Teaching and Learning)

Center for Machine Translation (CMT) / Carnegie Mellon University (CMU)

Centre d'expertise et de veille inforoutes et langues (CEVEIL)

COLING (International Conference on Computational Linguistics)

Comité européen pour le respect des cultures et des langues en Europe (CERCLE)

Computational Linguistics (CL) and Machine Translation (MT) Group (CL/MT Research Group) / Essex University

Computing Research Laboratory (CRL) / New Mexico State University (NMSU)

CTI (Computer in Teaching Initiative) Centre for Modern Languages / University of Hull

Dictionnaire francophone en ligne / Hachette & Agence universitaire de la Francophonie (AUPELF-UREF)

Dictionnaires électroniques / Administration fédérale suisse

ENGSPAN (SPANAM and ENGSPAN) / Organisation panaméricaine de la santé (PAHO)

Ethnologue (The)

EUROCALL (European Association for Computer-Assisted Language Learning)

Eurodicautom / Commission européenne

European Association for Machine Translation (EAMT)

European Chapter of the Association of Computational Linguistics (EACL)

European Minority Languages / Sabhal Mór Ostaig

European Network in Language and Speech (ELSNET)

FRANCIL (Réseau francophone de l'ingénierie de la langue) / Agence universitaire de la francophonie (AUPELF-UREF)

FRANTEXT / Institut national de la langue française (INaLF)

Global Reach

Globalink

Groupe d'étude pour la traduction automatique (GETA)

Human Language Technology (HLTCentral) / Commission européenne

Human-Languages Page (The)

ILOTERM / Organisation internationale du Travail (OIT)

Institut Dalle Molle pour les études sémantiques et cognitives (ISSCO)

Institut national de la langue française (INaLF)

International Committee on Computational Linguistics (ICCL)

International Conference on Computational Linguistics (COLING)

Internet Dictionary Project

Internet Resources for Language Teachers and Learners

Laboratoire de recherche appliquée en linguistique informatique (RALI)

Language Futures Europe

Language Today

Languages of the World by Computers and the Internet (The)

Lernout & Hauspie

LINGUIST List (The)

Localisation Industry Standards Association (LISA)

Logos (Canada, USA, Europe)

Logos (Italy)

Merriam-Webster Online: the Language Center

Multilingual Application Interface for Telematic Services (MAITS)

Multilingual Glossary of Internet Terminology (The) (Netglos) / WorldWide Language Institute (WWLI)

Multilingual Information Society (MLIS) / Commission européenne

Multilingual Tools and Services / Commission européenne

Natural Language Group (NLG) at USC/ISI / University of Southern California (USC)

NetGlos (The Multilingual Glossary of Internet Terminology) / WorldWide Language Institute (WWLI)

OneLook Dictionaries

Palmarès des langues de la Toile / Babel

PARC (Xerox Palo Alto Research Center)

Project Gutenberg

RALI (Laboratoire de recherche appliquée en linguistique informatique)

Réseau francophone de l'ingénierie de la langue (FRANCIL) / Agence universitaire de la francophonie (AUPELF-UREF)

SPANAM and ENGSPAN / Organisation panaméricaine de la santé (PAHO)

Speech on the Web

#### **SYSTRAN**

TERMITE (Base de données terminologique des Télécommunications de l'UIT) / Union internationale des télécommunications (UIT)

Théories et technologies multilingues (MLTT) / Xerox Research Centre Europe (XRCE)

Travlang

TTT.org (Translation, Theory and Technology) / Brigham Young University (BYU)

Unité de traduction assistée par ordinateur et de terminologie (CTT) / Organisation mondiale de la santé (OMS)

Universal Networking Language (UNL) / Université des Nations Unies (UNU)

W3C Internationalization/Localization / World Wide Web Consortium (W3C)

Web of Online Dictionaries (A)

WELL (Web Enhanced Language Learning)

WHOTERM (WHO Terminology Information System) / Organisation mondiale de la santé (OMS)

Xerox Palo Alto Research Center (PARC)

Xerox Research Centre Europe (XRCE)

Yamada WWW Language Guides

#### **8. INDEX DES PERSONNES CITEES**

L'astérisque (\*) signale les personnes qui ont particulièrement contribué à cette étude en participant à l'enquête menée par courrier électronique en juillet et décembre 1998.

Patrick Andries (Laboratoire de recherche appliquée en linguistique informatique - RALI)

Arlette Attali\* (Institut national de la langue française - INaLF)

Robert Beard\* (A Web of Online Dictionaries)

Louise Beaudoin (Ministère de la culture et des communications du Québec)

Guy Bertrand\* (Centre d'expertise et de veille inforoutes et langues - CEVEIL)

Christian Boitet\* (Groupe d'étude pour la traduction automatique - GETA)

Tyler Chambers\* (Human-Language Pages)

Jean-Pierre Cloutier (Chroniques de Cybérie)

Cynthia Delisle\* (Centre d'expertise et de veille inforoutes et langues - CEVEIL)

Helen Dry\* (LINGUIST List)

Bill Dunlap\* (2) (Euro-Marketing Associates, Global Reach)

Marcel Grangier\* (Section française des Services linguistiques centraux de la Chancellerie fédérale suisse)

Barbara F. Grimes\* (The Ethnologue)

Michael S. Hart\* (Project Gutenberg)

Randy Hobler\* (Globalink)

Eduard Hovy\* (Natural Language Group at USC/ISI)

Pierre Isabelle (Laboratoire de recherche appliquée en linguistique informatique - RALI)

Christiane Jadelot\* (Institut national de la langue française - INaLF)

Annie Kahn (Le Monde)

Brian King\* (NetGlos)

Geoffrey Kingscott\* (Praetorius)

Steven Krauwer\* (European Network in Language and Speech - ELSNET)

Michael C. Martin\* (Travlang)

Yoshi Mikami\* (The Languages of the World by Computer and the Internet)

Caoimhín P. Ó Donnáile\* (European Minority Languages)

Henri Slettenhaar\* (Webster University)

Martha L. Stone\* (2) (ZDNN)

June Thompson\* (CTI (Computer in Teaching Initiative) Centre for Modern Languages)

Paul Treanor\* (Language Futures Europe)

Rodrigo Vergara (Logos, Italie)

Robert Ware\* (2) (OneLook Dictionaries)

Copyright © 1999 Marie Lebert

End of Project Gutenberg's Le multilinguisme sur le Web, by Marie Lebert

\*\*\* END OF THE PROJECT GUTENBERG EBOOK LE MULTILINGUISME SUR LE WEB \*\*\*

Updated editions will replace the previous one—the old editions will be renamed.

Creating the works from print editions not protected by U.S. copyright law means that no one owns a United States copyright in these works, so the Foundation (and you!) can copy and distribute it in the United States without permission and without paying copyright royalties. Special rules, set forth in the General Terms of Use part of this license, apply to copying and distributing Project Gutenberg™ electronic works to protect the PROJECT GUTENBERG™ concept and trademark. Project Gutenberg is a registered trademark, and may not be used if you charge for an eBook, except by following the terms of the trademark license, including paying royalties for use of the Project Gutenberg trademark. If you do not charge anything for copies of this eBook, complying with the trademark license is very easy. You may use this eBook for nearly any purpose such as creation of derivative works, reports, performances and research. Project Gutenberg eBooks may be modified and printed and given away—you may do practically ANYTHING in the United States with eBooks not protected by U.S. copyright law. Redistribution is subject to the trademark license, especially commercial redistribution.

**START: FULL LICENSE**  
**THE FULL PROJECT GUTENBERG LICENSE**  
PLEASE READ THIS BEFORE YOU DISTRIBUTE OR USE THIS WORK

To protect the Project Gutenberg™ mission of promoting the free distribution of electronic works, by using or distributing this work (or any other work associated in any way with the phrase “Project Gutenberg”), you agree to comply with all the terms of the Full Project Gutenberg™ License available with this file or online at [www.gutenberg.org/license](http://www.gutenberg.org/license).

**Section 1. General Terms of Use and Redistributing Project Gutenberg™ electronic works**

1.A. By reading or using any part of this Project Gutenberg™ electronic work, you indicate that you have read, understand, agree to and accept all the terms of this license and intellectual property (trademark/copyright) agreement. If you do not agree to abide by all the terms of this agreement, you must cease using and return or destroy all copies of Project Gutenberg™ electronic works in your possession. If you paid a fee for obtaining a copy of or access to a Project Gutenberg™ electronic work and you do not agree to be bound by the terms of this agreement, you may obtain a refund from the person or entity to whom you paid the fee as set forth in paragraph 1.E.8.

1.B. “Project Gutenberg” is a registered trademark. It may only be used on or associated in any way with an electronic work by people who agree to be bound by the terms of this agreement. There are a few things that you can do with most Project Gutenberg™ electronic works even without complying with the full terms of this agreement. See paragraph 1.C below. There are a lot of things you can do with Project Gutenberg™ electronic works if you follow the terms of this agreement and help preserve free future access to Project Gutenberg™ electronic works. See paragraph 1.E below.

1.C. The Project Gutenberg Literary Archive Foundation (“the Foundation” or PGLAF), owns a compilation copyright in the collection of Project Gutenberg™ electronic works. Nearly all the individual works in the collection are in the public domain in the United States. If an individual work is unprotected by copyright law in the United States and you are located in the United States, we do not claim a right to prevent you from copying, distributing, performing, displaying or creating derivative works based on the work as long as all references to Project Gutenberg are removed. Of course, we hope that you will support the Project Gutenberg™ mission of promoting free access to electronic works by freely sharing Project Gutenberg™ works in compliance with the terms of this agreement for keeping the Project Gutenberg™ name associated with the work. You can easily comply with the terms of this agreement by keeping this work in the same format with its attached full Project Gutenberg™ License when you share it without charge with others.

This particular work is one of the few individual works protected by copyright law in the United States and most of the remainder of the world, included in the Project Gutenberg collection with the permission of the copyright holder. Information on the copyright owner for this particular work and the terms of use imposed by the copyright holder on this work are set forth at the beginning of this work.

1.D. The copyright laws of the place where you are located also govern what you can do with this work. Copyright laws in most countries are in a constant state of change. If you are outside the United States, check the laws of your country in addition to the terms of this agreement before downloading, copying, displaying, performing, distributing or creating derivative works based on this work or any other Project Gutenberg™ work. The Foundation makes no representations concerning the copyright status of any work in any country other than the United States.

1.E. Unless you have removed all references to Project Gutenberg:

1.E.1. The following sentence, with active links to, or other immediate access to, the full Project Gutenberg™ License must appear prominently whenever any copy of a Project Gutenberg™ work (any work on which the phrase “Project Gutenberg” appears, or with which the phrase “Project Gutenberg” is associated) is accessed, displayed, performed, viewed, copied or distributed:

This eBook is for the use of anyone anywhere in the United States and most other parts of the world at no cost and with almost no restrictions whatsoever. You may copy it, give it

away or re-use it under the terms of the Project Gutenberg License included with this eBook or online at [www.gutenberg.org](http://www.gutenberg.org). If you are not located in the United States, you will have to check the laws of the country where you are located before using this eBook.

1.E.2. If an individual Project Gutenberg™ electronic work is derived from texts not protected by U.S. copyright law (does not contain a notice indicating that it is posted with permission of the copyright holder), the work can be copied and distributed to anyone in the United States without paying any fees or charges. If you are redistributing or providing access to a work with the phrase "Project Gutenberg" associated with or appearing on the work, you must comply either with the requirements of paragraphs 1.E.1 through 1.E.7 or obtain permission for the use of the work and the Project Gutenberg™ trademark as set forth in paragraphs 1.E.8 or 1.E.9.

1.E.3. If an individual Project Gutenberg™ electronic work is posted with the permission of the copyright holder, your use and distribution must comply with both paragraphs 1.E.1 through 1.E.7 and any additional terms imposed by the copyright holder. Additional terms will be linked to the Project Gutenberg™ License for all works posted with the permission of the copyright holder found at the beginning of this work.

1.E.4. Do not unlink or detach or remove the full Project Gutenberg™ License terms from this work, or any files containing a part of this work or any other work associated with Project Gutenberg™.

1.E.5. Do not copy, display, perform, distribute or redistribute this electronic work, or any part of this electronic work, without prominently displaying the sentence set forth in paragraph 1.E.1 with active links or immediate access to the full terms of the Project Gutenberg™ License.

1.E.6. You may convert to and distribute this work in any binary, compressed, marked up, nonproprietary or proprietary form, including any word processing or hypertext form. However, if you provide access to or distribute copies of a Project Gutenberg™ work in a format other than "Plain Vanilla ASCII" or other format used in the official version posted on the official Project Gutenberg™ website ([www.gutenberg.org](http://www.gutenberg.org)), you must, at no additional cost, fee or expense to the user, provide a copy, a means of exporting a copy, or a means of obtaining a copy upon request, of the work in its original "Plain Vanilla ASCII" or other form. Any alternate format must include the full Project Gutenberg™ License as specified in paragraph 1.E.1.

1.E.7. Do not charge a fee for access to, viewing, displaying, performing, copying or distributing any Project Gutenberg™ works unless you comply with paragraph 1.E.8 or 1.E.9.

1.E.8. You may charge a reasonable fee for copies of or providing access to or distributing Project Gutenberg™ electronic works provided that:

- You pay a royalty fee of 20% of the gross profits you derive from the use of Project Gutenberg™ works calculated using the method you already use to calculate your applicable taxes. The fee is owed to the owner of the Project Gutenberg™ trademark, but he has agreed to donate royalties under this paragraph to the Project Gutenberg Literary Archive Foundation. Royalty payments must be paid within 60 days following each date on which you prepare (or are legally required to prepare) your periodic tax returns. Royalty payments should be clearly marked as such and sent to the Project Gutenberg Literary Archive Foundation at the address specified in Section 4, "Information about donations to the Project Gutenberg Literary Archive Foundation."
- You provide a full refund of any money paid by a user who notifies you in writing (or by e-mail) within 30 days of receipt that s/he does not agree to the terms of the full Project Gutenberg™ License. You must require such a user to return or destroy all copies of the works possessed in a physical medium and discontinue all use of and all access to other copies of Project Gutenberg™ works.
- You provide, in accordance with paragraph 1.F.3, a full refund of any money paid for a work or a replacement copy, if a defect in the electronic work is discovered and reported to you within 90 days of receipt of the work.
- You comply with all other terms of this agreement for free distribution of Project Gutenberg™ works.

1.E.9. If you wish to charge a fee or distribute a Project Gutenberg™ electronic work or group of works on different terms than are set forth in this agreement, you must obtain permission in writing from the Project Gutenberg Literary Archive Foundation, the manager of the Project Gutenberg™ trademark. Contact the Foundation as set forth in Section 3 below.

1.F.

1.F.1. Project Gutenberg volunteers and employees expend considerable effort to identify, do copyright research on, transcribe and proofread works not protected by U.S. copyright law in creating the Project Gutenberg™ collection. Despite these efforts, Project Gutenberg™ electronic works, and the medium on which they may be stored, may contain "Defects," such as, but not limited to, incomplete, inaccurate or corrupt data, transcription errors, a copyright or other intellectual property infringement, a defective or damaged disk or other medium, a computer virus, or computer codes that damage or cannot be read by your equipment.

1.F.2. LIMITED WARRANTY, DISCLAIMER OF DAMAGES - Except for the “Right of Replacement or Refund” described in paragraph 1.F.3, the Project Gutenberg Literary Archive Foundation, the owner of the Project Gutenberg™ trademark, and any other party distributing a Project Gutenberg™ electronic work under this agreement, disclaim all liability to you for damages, costs and expenses, including legal fees. YOU AGREE THAT YOU HAVE NO REMEDIES FOR NEGLIGENCE, STRICT LIABILITY, BREACH OF WARRANTY OR BREACH OF CONTRACT EXCEPT THOSE PROVIDED IN PARAGRAPH 1.F.3. YOU AGREE THAT THE FOUNDATION, THE TRADEMARK OWNER, AND ANY DISTRIBUTOR UNDER THIS AGREEMENT WILL NOT BE LIABLE TO YOU FOR ACTUAL, DIRECT, INDIRECT, CONSEQUENTIAL, PUNITIVE OR INCIDENTAL DAMAGES EVEN IF YOU GIVE NOTICE OF THE POSSIBILITY OF SUCH DAMAGE.

1.F.3. LIMITED RIGHT OF REPLACEMENT OR REFUND - If you discover a defect in this electronic work within 90 days of receiving it, you can receive a refund of the money (if any) you paid for it by sending a written explanation to the person you received the work from. If you received the work on a physical medium, you must return the medium with your written explanation. The person or entity that provided you with the defective work may elect to provide a replacement copy in lieu of a refund. If you received the work electronically, the person or entity providing it to you may choose to give you a second opportunity to receive the work electronically in lieu of a refund. If the second copy is also defective, you may demand a refund in writing without further opportunities to fix the problem.

1.F.4. Except for the limited right of replacement or refund set forth in paragraph 1.F.3, this work is provided to you ‘AS-IS’, WITH NO OTHER WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PURPOSE.

1.F.5. Some states do not allow disclaimers of certain implied warranties or the exclusion or limitation of certain types of damages. If any disclaimer or limitation set forth in this agreement violates the law of the state applicable to this agreement, the agreement shall be interpreted to make the maximum disclaimer or limitation permitted by the applicable state law. The invalidity or unenforceability of any provision of this agreement shall not void the remaining provisions.

1.F.6. INDEMNITY - You agree to indemnify and hold the Foundation, the trademark owner, any agent or employee of the Foundation, anyone providing copies of Project Gutenberg™ electronic works in accordance with this agreement, and any volunteers associated with the production, promotion and distribution of Project Gutenberg™ electronic works, harmless from all liability, costs and expenses, including legal fees, that arise directly or indirectly from any of the following which you do or cause to occur: (a) distribution of this or any Project Gutenberg™ work, (b) alteration, modification, or additions or deletions to any Project Gutenberg™ work, and (c) any Defect you cause.

## **Section 2. Information about the Mission of Project Gutenberg™**

Project Gutenberg™ is synonymous with the free distribution of electronic works in formats readable by the widest variety of computers including obsolete, old, middle-aged and new computers. It exists because of the efforts of hundreds of volunteers and donations from people in all walks of life.

Volunteers and financial support to provide volunteers with the assistance they need are critical to reaching Project Gutenberg™’s goals and ensuring that the Project Gutenberg™ collection will remain freely available for generations to come. In 2001, the Project Gutenberg Literary Archive Foundation was created to provide a secure and permanent future for Project Gutenberg™ and future generations. To learn more about the Project Gutenberg Literary Archive Foundation and how your efforts and donations can help, see Sections 3 and 4 and the Foundation information page at [www.gutenberg.org](http://www.gutenberg.org).

## **Section 3. Information about the Project Gutenberg Literary Archive Foundation**

The Project Gutenberg Literary Archive Foundation is a non-profit 501(c)(3) educational corporation organized under the laws of the state of Mississippi and granted tax exempt status by the Internal Revenue Service. The Foundation’s EIN or federal tax identification number is 64-6221541. Contributions to the Project Gutenberg Literary Archive Foundation are tax deductible to the full extent permitted by U.S. federal laws and your state’s laws.

The Foundation’s business office is located at 809 North 1500 West, Salt Lake City, UT 84116, (801) 596-1887. Email contact links and up to date contact information can be found at the Foundation’s website and official page at [www.gutenberg.org/contact](http://www.gutenberg.org/contact)

## **Section 4. Information about Donations to the Project Gutenberg Literary Archive Foundation**

Project Gutenberg™ depends upon and cannot survive without widespread public support and donations to carry out its mission of increasing the number of public domain and licensed works that can be freely distributed in machine-readable form accessible by the widest array of equipment

including outdated equipment. Many small donations (\$1 to \$5,000) are particularly important to maintaining tax exempt status with the IRS.

The Foundation is committed to complying with the laws regulating charities and charitable donations in all 50 states of the United States. Compliance requirements are not uniform and it takes a considerable effort, much paperwork and many fees to meet and keep up with these requirements. We do not solicit donations in locations where we have not received written confirmation of compliance. To SEND DONATIONS or determine the status of compliance for any particular state visit [www.gutenberg.org/donate](http://www.gutenberg.org/donate).

While we cannot and do not solicit contributions from states where we have not met the solicitation requirements, we know of no prohibition against accepting unsolicited donations from donors in such states who approach us with offers to donate.

International donations are gratefully accepted, but we cannot make any statements concerning tax treatment of donations received from outside the United States. U.S. laws alone swamp our small staff.

Please check the Project Gutenberg web pages for current donation methods and addresses. Donations are accepted in a number of other ways including checks, online payments and credit card donations. To donate, please visit: [www.gutenberg.org/donate](http://www.gutenberg.org/donate)

## **Section 5. General Information About Project Gutenberg™ electronic works**

Professor Michael S. Hart was the originator of the Project Gutenberg™ concept of a library of electronic works that could be freely shared with anyone. For forty years, he produced and distributed Project Gutenberg™ eBooks with only a loose network of volunteer support.

Project Gutenberg™ eBooks are often created from several printed editions, all of which are confirmed as not protected by copyright in the U.S. unless a copyright notice is included. Thus, we do not necessarily keep eBooks in compliance with any particular paper edition.

Most people start at our website which has the main PG search facility: [www.gutenberg.org](http://www.gutenberg.org).

This website includes information about Project Gutenberg™, including how to make donations to the Project Gutenberg Literary Archive Foundation, how to help produce our new eBooks, and how to subscribe to our email newsletter to hear about new eBooks.