

The Project Gutenberg eBook of Multilingualism on the Web

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

Title: Multilingualism on the Web

Author: Marie Lebert

Release date: October 26, 2008 [eBook #27028]

Language: English

*** START OF THE PROJECT GUTENBERG EBOOK MULTILINGUALISM ON THE WEB ***

Produced by Al Haines

MULTILINGUALISM ON THE WEB

MARIE LEBERT

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

Copyright © 1999 Marie Lebert

Dated February 1999, this study is divided into four parts: Multilingualism, Language Resources, Translation Resources and Language-Related Research. It is based on many interviews. With many thanks to Laurie Chamberlain, who kindly edited this paper. This study is also available in French: Le multilinguisme sur le web. The original versions are available on the NEF, University of Toronto: <http://www.etudes-francaises.net/entretiens/multi.htm>

TABLE OF CONTENTS

1. Introduction
2. Multilingualism
3. Language Resources
4. Translation Resources
5. Language-Related Research
6. Index of Websites
7. Index of Names

1. INTRODUCTION

It is true that the Internet transcends limitations of time, distances and borders, but what about languages?

From the beginning, the main language of the Internet has been English, and it still is today, but the use of other languages is steadily increasing. Sooner or later, the distribution of languages on the Internet will correspond to the language distribution on the planet, and free translation software in all languages will be available for an instantaneous translation of any website. But there is still a lot to do before multilingualism can be really effective.

This study is divided into four parts: Multilingualism; Language Resources; Translation Resources; and Language-Related Research.

In the chapter about multilingualism, we will study the growth of non-English languages on the Internet. French will be taken as an example, and the efforts in the European Union relating to the diversity of languages will be examined.

In the chapter about language resources, we will give some examples of the language resources available on the Web — sites indexing language resources, language directories, language dictionaries and glossaries, textual databases, and terminological databases.

In the chapter relating to translation resources, we will explore the problems and perspectives linked to machine translation and computer-assisted translation.

In the last chapter on language-related research, we will present some projects relating to machine translation research, computational linguistics, language engineering, and internationalization and localization.

In August and December 1998, I sent an inquiry, based on three questions, to organizations and companies involved in languages on the Web. The three questions were:

- a) How do you see multilingualism on the Internet?;
- b) What did the use of the Internet bring to your professional life and/or the life of your company/organization; and
- c) How do you see your professional future with the Internet or the future of Internet-related activities as regards languages?

The answers received are included in this study. I express here my warmest thanks to all those who sent me their comments.

[As a translator-editor - working mainly for the International Labour Office (ILO), Geneva, Switzerland - I am fascinated by languages in general, so I wanted to know more about multilingualism on the Web. I found I had some time to look into the subject and I wrote this paper about the topics I was particularly interested in (first version in November 1998, updated in February 1999). I am also interested in the relationship between the print media and the Internet, and I wrote another paper about these topics too.]

2. MULTILINGUALISM

[In this chapter:]

[2.1. The Web: First English, Then Multilingual / 2.2. A Non-English Language: The Example of French / 2.3. Diversity of Languages: The Situation in Europe]

2.1. The Web: First English, Then Multilingual

In the beginning, the Internet was nearly 100% English, which can be easily explained because it was created in the United States as a network set up by the Pentagon (in 1969) before spreading to US governmental agencies and to universities. After the creation of the World Wide Web in 1989-90 by Tim Berners-Lee at the European Laboratory for Particle Physics (CERN), in Geneva, Switzerland, and the distribution of the first browser Mosaic (the ancestor of Netscape) from November 1993 onwards, the Web too began to spread — first in the US thanks to considerable investments made by the government, then around North America, and then to the rest of the world.

The fact that there are many more Internet surfers in the US and Canada than in any other country is

due to different factors — these countries are among the leaders in the latest computing and communication technologies, and hardware and software, as well as local phone communications, are much cheaper there than in the rest of the world.

In Hugues Henry's article, *La francophonie en quête d'identité sur le Web*, published by the cybermagazine *Multimédium*, Jean-Pierre Cloutier, author of *Chroniques de Cybérie*, a weekly cybermagazine widely read in the French-speaking Internet community, explains:

"In Quebec I am spending about 120 hours per month on-line. My Internet access is \$30 [Canadian]; if I add my all-inclusive phone bill which is about \$40 (with various optional services), the total cost of my connection is \$70 per month. I leave you to guess what the price would be in France, in Belgium or in Switzerland, where the local communications are billed by the minute, for the same number of hours on-line."

It follows that Belgian, French or Swiss surfers spend much less time on the Web than they would like, or choose to surf at night to cut somehow their expenses.

In 1997, *Babel* — a joint initiative from Alis Technologies and the Internet Society, ran the first major study of the actual distribution of languages on the Internet. The results are published in the *Web Languages Hit Parade*, dated June 1997, and the languages, listed in order of usage, are: English 82.3%, German 4.0%, Japanese 1.6%, French 1.5%, Spanish 1.1%, Swedish 1.1%, and Italian 1.0%.

In *Web embraces language translation*, an article published in *ZDNN (ZD Network News)* of July 21, 1998, Martha L. Stone explained:

"This year, the number of new non-English websites is expected to outpace the growth of new sites in English, as the cyber world truly becomes a 'World Wide Web.' [...] According to Global Reach, the fastest growing groups of Web newbies are non-English-speaking: Spanish, 22.4 percent; Japanese, 12.3 percent; German, 14 percent; and French, 10 percent. An estimated 55.7 million people access the Web whose native language is not English. [...] Only 6 percent of the world population speaks English as a native language (16 percent speak Spanish), while about 80 percent of all web pages are in English."

According to Global Reach, 92% of the world does not speak English. As the Web quickly spreads worldwide, more and more operators of English-language sites which are concerned by the internationalization of the Web recognize that, although English may be the main international language for exchanges of all kinds, not everyone in the world reads English.

Since December 1997 any Internet surfer can use the AltaVista Translation service, which translates English web pages (up to three pages at the same time) into French, German, Italian, Portuguese, and Spanish, and vice versa. The Internet surfer can also buy and use Web translation software. In both cases he will get a usable but imperfect machine-translated result which may be very helpful, but will never have the same quality as a translation prepared by a human translator with special knowledge of the subject and the contents of the site.

The increase in multilingual sites will make it possible to include more diverse languages on the Internet. And more free translation software will improve communication among everyone in the international Internet community.

To reach as large an audience as possible, the solution is to create bilingual, trilingual, multilingual sites. The website of the Belgian daily newspaper *Le Soir* gives a presentation of the newspaper in six languages: French, English, Dutch, German, Italian and Spanish. The *French Club des poètes (Club of Poets)*, a French site dedicated to poetry, presents its site in English, Spanish and Portuguese. *E-Mail-Planet*, a free e-mail address provider, provides a menu in six languages (English, Finnish, French, Italian, Portuguese, and Spanish).

Robert Ware is the creator of *OneLook Dictionaries*, a fast finder for 2,058,544 words in 425 dictionaries in various fields: business, computer/Internet; medical; miscellaneous; religion; science; sports; technology; general; and slang. In his e-mail to me of September 2, 1998, he wrote:

"An interesting thing happened earlier in the history of the Internet and I think I learned something from it.

In 1994, I was working for a college and trying to install a software package on a particular type of computer. I located a person who was working on the same problem and we began exchanging e-mail. Suddenly, it hit me... the software was written only 30 miles away but I was getting help from a person half way around the world. Distance and geography no longer mattered!

OK, this is great! But what is it leading to? I am only able to communicate in English but, fortunately, the other person could use English as well as German which was his mother tongue. The Internet has removed one barrier (distance) but with that comes the barrier of language.

It seems that the Internet is moving people in two quite different directions at the same time. The Internet (initially based on English) is connecting people all around the world. This is further promoting a common language for people to use for communication. But it is also creating contact between people of different languages and creates a greater interest in multilingualism. A common language is great but in no way replaces this need.

So the Internet promotes both a common language AND multilingualism. The good news is that it helps provide solutions. The increased interest and need is creating incentives for people around the world to create improved language courses and other assistance and the Internet is providing fast and inexpensive opportunities to make them available."

2.2. A Non-English Language: The Example of French

Let us take French as an example of a non-English language.

Since 1996 the number of sites in French has increased significantly. There were about 20,000 sites in French in mid-1997, and more of a third of them were from Quebec. Since the beginning of 1998 we can see a larger number of new French websites, particularly in the field of electronic commerce. "For two years I have been waiting for France to wake up. Today I'll not complain about it," Louise Beaudouin, the Minister of Culture and Communications in Quebec, declared on February 10, 1998, when interviewed by the daily cybermagazine Multimédium.

Until early 1998, Quebec and its 6 million inhabitants had more websites than France did with its 60 million inhabitants. In her interview, Louise Beaudouin gave two reasons for France's lagging behind Quebec — the first is the high cost of phone service, and the second is the widespread use of the Minitel for commercial transactions.

Developed 15 years ago by France Télécom, the French state telephone company, the Minitel is a terminal which gives access to the French videotex network, as well as facilitating electronic commerce transactions. As this very handy tool has been in use for years, it slowed down the expansion of French electronic commerce on the Internet. Little by little, many of the French companies or organizations with Minitel servers are creating websites, which are cheaper to consult, easier to use because of hypertext links, and more pleasing to the eye because of colors, graphics and multimedia tools.

French is not only spoken in France, Quebec, and parts of Belgium and Switzerland, it is the official language of 49 states (particularly in Africa) and is spoken worldwide by 500 million people. Created in 1970 with 21 French-speaking states, the Agence de la francophonie (Agency of Francophone Countries) counts 47 members today. Its goal is to be an instrument of multilateral cooperation to create a community representing the French-speaking countries at the international level.

Following the decisions of the Heads of States and Governments of French-speaking Countries during their meeting in Hanoi, Vietnam, in November 1997, the Fonds francophone des inforoutes (Francophone Fund for Information Highways) was established on June 3, 1998. Thirteen Francophone states and governments participated: the Belgian-French Community, Benin, Cameroon, Canada, Canada-New Brunswick, Canada-Quebec, Côte d'Ivoire, France, Gabon, Lebanon, Monaco, Senegal, and Switzerland.

This Fund's mission had been outlined six months earlier, according to several directives given by the Conférence des ministres chargés des inforoutes (Conference of Ministers in Charge of the Information Highways) held in Montreal, Quebec, in May 1997. It supported: democratization of the access to information highways; development of education, training and research; reinforcement of content creation and circulation; promotion of economic and social development; setting up of a Francophone awareness service; awareness-raising of young people, producers and investors; setting up of a concerted Francophone presence within the international authorities in charge of the development of information highways. The Fund's activities are particularly aimed at financing multilateral projects which would strengthen partnerships between North and South.

French is not only the language of 49 countries and 500 million inhabitants in the world, it is also the second international language used in international organizations. Despite the real and alleged pressure of the English-speaking community, French-speaking people insist on their language being given a fair position in the world, and receiving the same consideration given to other main languages of communication, such as English, Arabic, Chinese or Spanish. Just as for any other non-English

language-based culture, the French wish to stand up for their own language as well as for multilingualism and the diversity of people and culture.

At present it is important for any language to be represented through websites in its own language, with the possibility for Internet surfers to study it in a dynamic way through self-taught programs, language dictionaries, or linguistic databases. For example, in France, the Institut national de la langue française (INaLF) (National Institute of the French Language) created its site in December 1997 to present its research programs on the French language, particularly its lexicon. The INaLF's constantly expanded and renewed data, processed by specific and original computing systems, deal with all the aspects of the French language: literary discourse (14th-20th centuries), standard language (written and spoken), scientific and technical language (terminologies), and regional languages.

In her e-mail response of June 8, 1998, Christiane Jadelot, an engineer at INaLF-Nancy, France, explained:

"At the request of Robert Martin, the Head of INaLF, our first pages were posted on the Internet by mid-1996. I participated in the creation of these web pages with tools that cannot be compared to the ones we have nowadays. I was working with tools on UNIX, which were not very easy to use. At this time, we had little experience in this field, and the pages were very wordy. But the managing team was thinking it was urgent for us to be known through the Internet, a tool many enterprises were already using to promote their products. As we are a Department of Research and Services (Unité de recherche et de service), we have to find clients for our computer products, the best known being the textual database FRANTEXT. I think FRANTEXT was already on the Internet [since early 1995], and there was also a prototype of the volume 14 of the TLF [Trésor de la langue française (Treasure of the French Language), by Jean Nicot, 1606]. Therefore it was necessary for INaLF activities to be known by this means. It corresponded to a general need."

Every non-English language community is working for its language to be represented on the Web and for the international Internet to be multilingual. As an example, a non-profit organization created by the Government of Quebec, the Centre d'expertise et de veille Inforoutes et Langues (CEVEIL) (Centre of Expertise and Awareness for Information Highways and Languages) is setting up, in a more specifically French-oriented approach, an expertise network and some awareness-raising activities on the language problems of information highways.

Guy Bertrand, scientific director of CEVEIL, and Cynthia Delisle, consultant, answered my questions in their e-mail of August 23, 1998.

ML: "How do you see multilingualism on the Web?"

CEVEIL: "Multilingualism on the Internet is the logical and natural consequence of the diversity of human populations. Because the Web has first been developed and used in the United States, it is not really surprising that this medium began by being essentially Anglophone (and still is at present). However this situation is beginning to change and this movement will go on expanding, both because most of the new network users will not have English as a mother tongue and because the [non-English] communities already present on the Web will no longer accept the hegemony of the English language and will want to use the Internet in their own language, at least partially.

We can plan that, in several years, we'll have a situation similar to the one in publishing regarding the representation of different languages. This means that only a small number of languages will be in use (compared to the several thousands which exist). In this perspective, we believe that the Web — among other parties — should seek to further support minority cultures and languages, particularly for dispersed communities.

Finally, the arrival on the Internet of languages other than English, while requiring true readjustments and providing undeniable enrichment, points out the need for linguistic processing tools capable of effectively managing this situation. These will emerge as the result of research studies and awareness activities in areas such as machine translation, standardization, information location, automatic condensation (summaries), etc."

ML: "What did the use of the Internet bring to the life of CEVEIL?"

CEVEIL: "Let us first mention that the existence of the Web is one of the grounds of existence of CEVEIL, as we concentrate our activities mainly around the set of themes of the language use and processing on the Internet.

Moreover the Web is our main field for gathering information on the set of themes we are concerned with. Among others, we regularly and frequently watch the sites circulating daily and/or weekly news. At this level, we can say without hesitation that we use the Internet more than the other available

written resources to carry out our activities.

Otherwise we prolifically use electronic mail to maintain relations with our contributors in order to obtain information and realize some projects. CEVEIL is a 'network structure' which would survive with difficulty without the Internet to connect together all the people who are implicated.

Finally it is useful to point out that the Web is also our most important tool for distributing our products to our target clients: sending of electronic news reports to our subscribers, creation of an electronic periodical, information and document distribution via our website, etc."

ML: "How does CEVEIL see the future of Internet-related activities as regards languages?"

CEVEIL: "The Internet is here to stay. The arrival of languages other than English to this medium also is irreversible. Therefore it is necessary to take these new facts into consideration from an economic, social, political, cultural, etc., point of view. Sectors such as advertising, vocational training, work in groups or within networks and knowledge management, will consequently have to evolve. As we mentioned above, it brings us back to the necessary development of really effective technologies and tools which will further exchanges in a really multilingual global village..."

2.3. Diversity of Languages: The Situation in Europe

Henri Slettenhaar, professor at the Webster University, Geneva, Switzerland, is a trilingual European. He is Dutch, he teaches computer science in English, and he speaks French too because he lives in France. He answered my questions in his e-mail of December 21, 1998.

ML: "How do you see multilingualism on the Internet?"

HS: "I see multilingualism as a very important issue. Local communities which are on the Web should use the local language first and foremost for their information. If they want to be able to present their information to the world community as well, their information should be in English as well. I see a real need for bilingual websites."

ML: "How do you see the future of Internet-related activities as regards languages?"

HS: "As far as languages are concerned, I am delighted that there are so many offerings in the original languages now. I much prefer to read the original with difficulty than to get a bad translation."

According to Global Reach, only 15% of Europe's half a billion population speaks English as a first language, and only 28% speaks English at all. A recent study showed that only 32% of Web surfers on the European continent consult the Web in English.

Founder of Euro-Marketing Associates (including Global Reach), Bill Dunlap, who champions European e-commerce among his fellow American compatriots, explained in his e-mail of December 12, 1998 that, contrary to North America, "in Europe [...], the countries are small enough so that an international perspective has been necessary for centuries."

There are many European organizations dealing with multilingualism, such as the European Language Resources Association (ELRA), the European Network in Language and Speech (ELSNET) and the Multilingual Information Society (MLIS) Programme of the European Union.

The European Language Resources Association (ELRA) was established as a non-profit organization in Luxembourg in February 1995. Its overall goal is to provide a centralized organization for the validation, management, and distribution of speech, text, and terminology resources and tools, and to promote their use within the European telematics RTD (research and technological development) community. Its website is bilingual English-French.

The European Network in Language and Speech (ELSNET) has over a hundred European academic and industrial institutions as members. The long-term technological goal which unites the participants of ELSNET is to build multilingual speech and NL (natural language) systems with unrestricted coverage of both spoken and written language.

In his e-mail of September 23, 1998, Steven Krauwer, ELSNET coordinator, explained:

"— as a European citizen I think that multilingualism on the Web is absolutely essential, as in the long run I don't think that it is a healthy situation when only those who have a reasonable command of English can fully exploit the benefits of the Web;

— as a researcher (specialized in machine translation) I see multilingualism as a major challenge: how can we ensure that all information on the Web is accessible to everybody, irrespective of language differences.

[The Internet] is my main instrument to communicate with others, and it is my main source of information. [...] I am sure I will spend the rest of my professional life trying to use IT to take away or at least lower the language barriers."

The Multilingual Information Society (MLIS) Programme of the European Union promotes the linguistic diversity of the EU in the information society. It intends to raise awareness of and stimulate provision of multilingual services, tolerable conditions for the language industries, reduced cost of information transfer among languages and contribute to the promotion of linguistic diversity. The home page of the website is in English, and documents are issued in many of all 11 EU official languages: Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, and Swedish.

Linguistic pluralism and diversity are everybody's business, as explained in a petition launched by the European Committee for the Respect of Cultures and Languages in Europe (ECRCLE) "for a humanist and multilingual Europe, rich of its cultural diversity".

"Linguistic pluralism and diversity are not obstacles to the free circulation of men, ideas, goods and services, as would like to suggest some objective allies, consciously or not, of the dominant language and culture. Indeed, standardization and hegemony are the obstacles to the free blossoming of individuals, societies and the information economy, the main source of tomorrow's jobs. On the contrary, the respect for languages is the last hope for Europe to get closer to the citizens, an objective always claimed and almost never put into practice. The Union must therefore give up privileging the language of one group."

The full text of the petition is available on the Web in the 11 European official languages of the European Union. The ECRCLE also asks the revisors of the Treaty of the European Union to include in the text of the treaty the respect of national cultures and languages. The proposals are concrete. In particular, the petition asks the governments in each country to "teach the youth at least two, and preferably three foreign European languages; encourage the national audiovisual and musical industries; and favour the diffusion of European works."

In Language Futures Europe, Paul Treanor collects links on language policy, multilingualism, global language structures, and the dominance of English. The site starts with a comment on the structures of language. It offers texts and essays, sections on EU policy, national policies, and research sites, and links on the emerging "monolingual movement" in the United States.

In his e-mail of August 18, 1998, Paul Treanor sent his comments on the questions I sent him:

"First, you speak of the Web in the singular. As you may have read, I think 'THE WEB' is a political, not a technological concept. A civilization is possible with extremely advanced computers, but no interconnection. The idea that there should be ONE WEB is derived from the liberal tradition of the single open, preferably global market.

I already suggested that the Internet should simply be broken up, and that Europe should cut the links with the US, and build a systematically incompatible net for Europe. As soon as you imagine the possibility of multiple nets, the language issues you list in your study are often irrelevant. Remember that 15 years ago, everyone thought that there would be one global TV station, CNN. Now there are French, German, Spanish global TV channels. So the answer to your question is that the 'one web' will split up anyway: probably into these 4 components:

- a) an internal US/Canadian anglophone net, with many of the original characteristics;
- b) separate national nets, with limited outside links;
- c) a new global net specifically to link the nets of category 2;
- d) possibly a specific EU net.

As you see, this structure parallels the existing geopolitical structure. All telecommunications infrastructure has followed similar patterns.

I think that it is not possible to approach the Web in the neutral apolitical way suggested by your study. Current EU policy pretends to be neutral in this way, but in fact is supporting the growth of English as a contact-language in EU communications policy."

3. LANGUAGE RESOURCES

[In this chapter:]

[3.1. Sites Indexing Language Resources / 3.2. Language Directories / 3.3. Dictionaries and Glossaries / 3.4. Textual Databases / 3.5. Terminological Databases]

3.1. Sites Indexing Language Resources

Prepared by the Telematics for Libraries Programme of the European Union, Multilingual Tools and Services gives a series of links to dictionaries, multilingual support, projects, search engines by language, terminology data banks, thesauri, and translation systems.

Created by Tyler Chambers in May 1994, The Human-Languages Page is a comprehensive catalog of 1,800 language-related Internet resources in more than 100 different languages. The subject listings are: languages and literature; schools and institutions; linguistics resources; products and services; organizations; jobs and internships. The category listings are: dictionaries and language lessons.

Tyler Chambers' other main language-related project is the Internet Dictionary Project. As explained on the website:

"The Internet Dictionary Project's goal is to create royalty-free translating dictionaries through the help of the Internet's citizens. This site allows individuals from all over the world to visit and assist in the translation of English words into other languages. The resulting lists of English words and their translated counterparts are then made available through this site to anyone, with no restrictions on their use. [...]

The Internet Dictionary Project began in 1995 in an effort to provide a noticeably lacking resource to the Internet community and to computing in general — free translating dictionaries. Not only is it helpful to the on-line community to have access to dictionary searches at their fingertips via the World Wide Web, it also sponsors the growth of computer software which can benefit from such dictionaries — from translating programs to spelling-checkers to language-education guides and more. By facilitating the creation of these dictionaries on-line by thousands of anonymous volunteers all over the Internet, and by providing the results free-of-charge to anyone, the Internet Dictionary Project hopes to leave its mark on the Internet and to inspire others to create projects which will benefit more than a corporation's gross income."

Tyler Chambers answered my questions in his e-mail of 14 September 1998.

ML: "How do you see multilingualism on the Web?"

TC: "Multilingualism on the Web was inevitable even before the medium 'took off', so to speak. 1994 was the year I was really introduced to the Web, which was a little while after its christening but long before it was mainstream. That was also the year I began my first multilingual Web project, and there was already a significant number of language-related resources on-line. This was back before Netscape even existed — Mosaic was almost the only Web browser, and web pages were little more than hyperlinked text documents. As browsers and users mature, I don't think there will be any currently spoken language that won't have a niche on the Web, from Native American languages to Middle Eastern dialects, as well as a plethora of 'dead' languages that will have a chance to find a new audience with scholars and others alike on-line. To my knowledge, there are very few language types which are not currently on-line: browsers currently have the capability to display Roman characters, Asian languages, the Cyrillic alphabet, Greek, Turkish, and more. Accent Software has a product called 'Internet with an Accent' which claims to be able to display over 30 different language encodings. If there are currently any barriers to any particular language being on the Web, they won't last long."

ML: "What did the use of the Internet bring to your professional life?"

TC: "My professional life is currently completely separate from my Internet life. Professionally, I'm a computer programmer/techie — I find it challenging and it pays the bills. On-line, my work has been with making language information available to more people through a couple of my Web-based projects. While I'm not multilingual, nor even bilingual, myself, I see an importance to language and multilingualism that I see in very few other areas. The Internet has allowed me to reach millions of people and help them find what they're looking for, something I'm glad to do. It has also made me somewhat of a celebrity, or at least a familiar name in certain circles — I just found out that one of my Web projects had a short mention in Time Magazine's Asia and International issues. Overall, I think that the Web has been great for language awareness and cultural issues — where else can you

randomly browse for 20 minutes and run across three or more different languages with information you might potentially want to know? Communications mediums make the world smaller by bringing people closer together; I think that the Web is the first (of mail, telegraph, telephone, radio, TV) to really cross national and cultural borders for the average person. Israel isn't thousands of miles away anymore, it's a few clicks away — our world may now be small enough to fit inside a computer screen."

ML: "How do you see the future of Internet-related activities as regards languages?"

TC: "As I've said before, I think that the future of the Internet is even more multilingualism and cross-cultural exploration and understanding than we've already seen. But the Internet will only be the medium by which this information is carried; like the paper on which a book is written, the Internet itself adds very little to the content of information, but adds tremendously to its value in its ability to communicate that information. To say that the Internet is spurring multilingualism is a bit of a misconception, in my opinion — it is communication that is spurring multilingualism and cross-cultural exchange, the Internet is only the latest mode of communication which has made its way down to the (more-or-less) common person. The Internet has a long way to go before being ubiquitous around the world, but it, or some related progeny, likely will. Language will become even more important than it already is when the entire planet can communicate with everyone else (via the Web, chat, games, e-mail, and whatever future applications haven't even been invented yet), but I don't know if this will lead to stronger language ties, or a consolidation of languages until only a few, or even just one remain. One thing I think is certain is that the Internet will forever be a record of our diversity, including language diversity, even if that diversity fades away. And that's one of the things I love about the Internet — it's a global model of the saying 'it's not really gone as long as someone remembers it'. And people do remember."

Since its inception in 1989, the CTI (Computer in Teaching Initiative) Centre for Modern Languages has been based in the Language Institute at the University of Hull, United Kingdom, and aims to promote and encourage the use of computers in language learning and teaching. The Centre provides information on how computer assisted language learning (CALL) can be effectively integrated into existing courses and offers support for language lecturers who are using, or who wish to use, computers in their teaching.

June Thompson, Manager of the Centre, answered my questions in his e-mail of December 14, 1998.

ML: "How do you see multilingualism on the Internet?"

JT: "The Internet has the potential to increase the use of foreign languages, and our organisation certainly opposed any trend towards the dominance of English as the language of the Internet. An interesting paper on this topic was delivered by Madanmohan Rao at the WorldCALL conference in Melbourne, July 1998." [See details of the forthcoming conference book]

ML: "What did the use of the Internet bring to the life of your organization?"

JT: "The use of the Internet has brought an enormous new dimension to our work of supporting language teachers in their use of technology in teaching."

ML: "How do you see the future of Internet-related activities as regards languages?"

JT: "I suspect that for some time to come, the use of Internet-related activities for languages will continue to develop alongside other technology-related activities (e.g. use of CD-ROMs - not all institutions have enough networked hardware). In the future I can envisage use of Internet playing a much larger part, but only if such activities are pedagogy-driven. Our organisation is closely associated with the WELL project [Web Enhanced Language Learning] which devotes itself to these issues."

Hosted by the CTI Centre for Modern Languages and the University of Hull (United Kingdom), EUROCALL is the European Association for Computer Assisted Language Learning. This association of language teaching professionals from Europe and worldwide aims to: promote the use of foreign languages within Europe; provide a European focus for all aspects of the use of technology for language learning; enhance the quality, dissemination and efficiency of CALL (computer assisted language learning) materials; and support Special Interest Groups (SIGs): CAPITAL (Computer Assisted Pronunciation Investigation Teaching and Learning), a group of researchers and practitioners interested in using the computers in the domain of pronunciation in the widest sense of the word, and WELL (Web Enhanced Language Learning), which will provide access to high-quality Web resources in 12 languages, selected and described by subject experts, plus information and examples on how to use them for teaching and learning.

Internet Resources for Language Teachers and Learners offers several categories of links: general

languages resources (centres and departments, dictionaries and grammars; discussion lists; distance language learning; fonts; journals; linguistics; lists and indexes; miscellaneous; newspapers and periodicals; organizations; resource sites; software; translation and interpreting); language-specific resources; multilingual language sites; search engines and indexes; and commercial language sites (audiovisual, language schools, resources and directories, software).

Maintained by the Institute of Phonetic Sciences, Amsterdam, the Netherlands, Speech on the Web is an extensive list of links organized in various sections: congresses, meetings, and workshops; links and lists; phonetics and speech; natural language processing, cognitive science, and AI (artificial intelligence); computational linguistics; dictionaries; electronic newsletters, journals and publications.

Travlang is a site dedicated both to travel and languages. Created by Michael C. Martin in 1994 on the site of his university when he was a student in physics, Foreign Languages for Travelers, included in Travlang in 1995, gives the possibility to learn 60 different languages on the Web. Translating Dictionaries gives access to free dictionaries in various languages (Afrikaans, Czech, Danish, Dutch, Esperanto, Finnish, French, Frisian, German, Hungarian, Italian, Latin, Norwegian, Portuguese, and Spanish). Maintained by its founder, who is now a researcher in experimental physics at the Lawrence Berkeley National Laboratory, California, the site offers numerous links to language dictionaries, translation services, language schools, multilingual bookstores, etc.

Michael C. Martin answered my questions in his e-mail of August 25, 1998.

ML: "How do you see multilingualism on the Web?"

MCM: "I think the Web is an ideal place to bring different cultures and people together, and that includes being multilingual. Our Travlang site is so popular because of this, and people desire to feel in touch with other parts of the world."

ML: "What did the use of the Internet bring to your professional life?"

MCM: "Well, certainly we've made a little business of it! The Internet is really a great tool for communicating with people you wouldn't have the opportunity to interact with otherwise. I truly enjoy the global collaboration that has made our Foreign Languages for Travelers pages possible."

ML: "How do you see the future of Internet-related activities as regards languages?"

MCM: "I think computerized full-text translations will become more common, enabling a lot of basic communications with even more people. This will also help bring the Internet more completely to the non-English speaking world."

The LINGUIST List is the component of the WWW Virtual Library for linguistics. It gives an extensive series of links on linguistic resources: the profession (conferences, linguistic associations, programs, etc.); research and research support (papers, dissertation abstracts, projects, bibliographies, topics, texts); publications; pedagogy; language resources (languages, language families, dictionaries, regional information); and computer support (fonts and software).

Helen Dry, moderator of the LINGUIST List, explained in her e-mail of August 18, 1998:

"The LINGUIST List, which I moderate, has a policy of posting in any language, since it's a list for linguists. However, we discourage posting the same message in several languages, simply because of the burden extra messages put on our editorial staff. (We are not a bounce-back list, but a moderated one. So each message is organized into an issue with like messages by our student editors before it is posted.) Our experience has been that almost everyone chooses to post in English. But we do link to a translation facility that will present our pages in any of 5 languages; so a subscriber need not read LINGUIST in English unless s/he wishes to. We also try to have at least one student editor who is genuinely multilingual, so that readers can correspond with us in languages other than English."

Maintained by the Yamada Language Center of the University of Oregon, the Yamada WWW Language Guides is a directory of language resources by geographic family and alphabetic family. It covers organizations, teaching institutes, curriculum materials, cultural references, and WWW links.

Language today is a new magazine for people working in applied languages: translators, interpreters, terminologists, lexicographers and technical writers. It is a collaborative project between Logos, who provide the website, and Praetorius, the UK language consultancy which keeps itself constantly informed about developments in applied languages. The site gives links to translators associations, language schools, and dictionaries.

Geoffrey Kingscott, managing director of Praetorius, answered my questions in his e-mail of

September 4, 1998.

ML: "How do you see multilingualism on the Web?"

GK: "Because the salient characteristics of the Web are the multiplicity of site generators and the cheapness of message generation, as the Web matures it will in fact promote multilingualism. The fact that the Web originated in the USA means that it is still predominantly in English but this is only a temporary phenomenon. If I may explain this further, when we relied on the print and audiovisual (film, television, radio, video, cassettes) media, we had to depend on the information or entertainment we wanted to receive being brought to us by agents (publishers, television and radio stations, cassette and video producers) who have to subsist in a commercial world or — as in the case of public service broadcasting — under severe budgetary restraints. That means that the size of the customer-base is all-important, and determines the degree to which languages other than the ubiquitous English can be accommodated. These constraints disappear with the Web. To give only a minor example from our own experience, we publish the print version of Language Today only in English, the common denominator of our readers. When we use an article which was originally in a language other than English, or report an interview which was conducted in a language other than English, we translate into English and publish only the English version. This is because the number of pages we can print is constrained, governed by our customer-base (advertisers and subscribers). But for our Web edition we also give the original version."

ML: "What did the use of the Internet bring to your company?"

GK: "The Internet has made comparatively little difference to our company. It is an additional medium rather than one which will replace all others."

ML: "How do you see the future with the Internet?"

GK: "We will continue to have a company website, and to publish a version of the magazine on the Web, but it will remain only one factor in our work. We do use the Internet as a source of information which we then distill for our readers, who would otherwise be faced with the biggest problem of the Web — indiscriminating floods of information."

3.2. Language Directories

The Ethnologue is the electronic version of The Ethnologue, 13th ed., (editor: Barbara F. Grimes, consulting editors: Richard S. Pittman and Joseph E. Grimes), published in 1996 by the Summer Institute of Linguistics, Dallas, Texas. This catalogue of more than 6,700 languages spoken in 228 countries is accessible through two search tools: The Ethnologue Name Index, which lists language names, dialect names, and alternate names, and The Ethnologue Language Family Index, which organizes languages according to language families.

Barbara F. Grimes, editor of The Ethnologue, wrote in her e-mail of August 18, 1998:

"Multilingual web pages are more widely useful, but much more costly to maintain. We have had requests for The Ethnologue in a few other languages, but we do not have the personnel or funds to do the translation or maintenance, since it is constantly being updated.

We have found the Internet to be useful, convenient, and supplementary to our work. Our main use of it is for e-mail.

It is a convenient means of making information more widely available to a wider audience than the printed Ethnologue provides.

On the other hand, many people in the audience we wish to reach do not have access to computers, so in some ways the Ethnologue on Internet reaches a limited audience who own computers. I am particularly thinking of people in the so-called 'third world'."

Created in December 1995 by Yoshi Mikami of Asia Info Network, The Languages of the World by Computers and the Internet (commonly called Logos Home Page or Kotoba Home Page) gives, for each language, its brief history, features, writing system, and character set and keyboard for computers and the Internet processing. In his e-mail of December 17, 1998, Yoshi Mikami wrote:

"My native tongue is Japanese. Because I had my graduate education in the US and worked in the computer business, I became bilingual Japanese/American English. I was always interested in different languages and cultures, so I learned some Russian, French and Chinese along the way. In late 1995, I created on the Web The Languages of the World by Computers and the Internet and tried to summarize

there the brief history, linguistic and phonetic features, writing system and computer processing for each of the six major languages of the world, in English and Japanese. As I gained more experience, I invited my two associates to write a book on viewing, understanding and creating the multilingual web pages, which was published in August, 1997, as "The Multilingual Web Guide" (see its support page) in the Japanese edition, the world's first book on such a subject.

Thousands of years ago, in Egypt, China and elsewhere, people were more conscious about communicating their laws and thoughts not in just one language, but in different languages. In our modern world, each nation state has adopted more or less one language for its own use. I see in the future of the Internet a greater use of different languages and multilingual pages, not a simple gravitation to American English, and a more creative use of multilingual computer translation. Ninety nine percent of the Webs created in Japan are written in Japanese!"

Maintained on the website of the College Sabhal Mór Ostaig, Island of Skye, Scotland, by Caoimhín P. Ó Donnáile, European Minority Languages is a list of minority languages by alphabetic order and by language family. The site also gives links to other sites dealing with the same subject worldwide.

Caoimhín P. Ó Donnáile wrote in her e-mail of August 18, 1998:

"— The Internet has contributed and will contribute to the wildfire spread of English as a world language.

— The Internet can greatly help minority languages, but this will not happen by itself. It will only happen if people want to maintain the language as an aim in itself.

— The Web is very useful for delivering language lessons, and there is a big demand for this.

— The Unicode (ISO 10646) character set standard is very important and will greatly assist in making the Internet more multilingual."

3.3. Dictionaries and Glossaries

There are more and more on-line dictionaries. Let us give three examples (English, French and multilingual).

In Merriam-Webster Online: the Language Center, a main publisher of English dictionaries gives free access to a collection of on-line resources. The goal is to help track down definitions, spellings, pronunciations, synonyms, vocabulary exercises, and other key facts about words and language. The main on-line resources are: WWWebster Dictionary, WWebster Thesaurus, Webster's Third (a lexical landmark), Guide to International Business Communications, Vocabulary Builder (with interactive vocabulary quizzes), and the Barnhart Dictionary Companion (hot new words).

The Dictionnaire francophone en ligne is the web version of the Dictionnaire universel francophone, published by Hachette, a major French publisher, and the Agence universitaire de la Francophonie (AUPELF-UREF) (University Agency for Francophony), which presents the standard French and the French words and expressions used in the five continents.

The Logos Dictionary is a multilingual dictionary with 8 million entry words in all languages. Logos, an international translation company based in Modena, Italy, gives free access to the linguistic tools used by its translators: 200 translators in its headquarters and 2,500 translators on-line all over the world, who process around 200 texts per day. Apart from the Logos Dictionary, these tools include: the Wordtheque, a word-by-word multilingual library with a massive database (325 million words) containing multilingual novels, technical literature and translated texts; Linguistic Resources, a database of 536 glossaries; and the Universal Conjugator, a database for conjugation of verbs in 17 languages.

In *Les mots pour le dire*, an article of the French daily newspaper *Le Monde* of December 7, 1997, Annie Kahn wrote:

"The Logos site is much more than a mere dictionary or a collection of links to other on-line dictionaries. A cornerstone of the system is the document search software, which processes a corpus of literary texts available free of charge on the Web. If you search for the definition or the translation of a word ('didactique', for example), you get not only the answer sought, but also a quote from one of the literary works containing the word (in our case, an essay by Voltaire). All it takes is a click on the mouse to access the whole text or even to order the book, thanks to a partnership agreement with Amazon.com, the well-known on-line book shop. Foreign translations are also available. If however no text containing the required word is found, the system acts as a search engine, sending the user to

other websites concerning the term in question. In the case of certain words, you can even hear the pronunciation. If there is no translation currently available, the system calls on the public to contribute. Everyone can make their own suggestion, after which Logos translators and the company verify the translations forwarded."

In the same article, Rodrigo Vergara, the Head of Logos, explained:

"We wanted all our translators to have access to the same translation tools. So we made them available on the Internet, and while we were at it we decided to make the site open to the public. This made us extremely popular, and also gave us a lot of exposure. In fact the operation attracted a great number of customers, and also allowed us to widen our network of translators, thanks to the contacts made in the wake of this initiative."

The dictionary directories are invaluable tools for linguists, such as Dictionnaires électroniques (Electronic Dictionaries), OneLook Dictionaries and A Web of Online Dictionaries.

Dictionnaires électroniques (Electronic Dictionaries) is an extensive list of electronic dictionaries prepared by the Section française des Services linguistiques centraux (SLC-f) (French Section of the Central Linguistic Services) of the Swiss Federal Administration, and classified into five main sections: abbreviations and acronyms; monolingual dictionaries; bilingual dictionaries; multilingual dictionaries; and geographical information. The search of a dictionary is also possible by key-words.

Marcel Grangier, head of this section, answered my questions in his e-mail of January 14, 1999.

ML: "How do you see multilingualism on the Internet?"

MG: "Multilingualism on the Internet can be seen as a happy and above all irreversible inevitability. In this perspective we have to make fun of the wet blankets who only speak to complain about the supremacy of English. This supremacy is not wrong in itself, inasmuch as it is the result of mainly statistical facts (more PCs per inhabitant, more English-speaking people, etc.). The counter-attack is not to 'fight against English' and even less to whine about it, but to increase sites in other languages. As a translation service, we also recommend the multilingualism of websites."

ML: "What did the use of the Internet bring to your professional life?"

MG: "To work without the Internet is simply impossible now — as well as all the tools used (e-mail, electronic press, services for translators), Internet is for us an essential and inexhaustible source of information in what I would call the 'non-structured sector' of the Web. For example, when the answer to a translation problem can't be found in websites presenting information in an organized way, in most cases search engines allow us to find the missing link somewhere on the network."

ML: "How do you see the future of Internet-related activities as regards languages?"

MG: "The increase in the number of languages on the Internet is inevitable, and can only be a benefit for multicultural exchanges. For the exchanges to happen in an optimal environment, it is still necessary to develop tools which will improve compatibility — the complete management of diacritics is only one example of what can be done."

Provided as a free service since April 1996 by Study Technologies, Englewood, Colorado, OneLook Dictionaries, by Robert Ware, is the fastest finder for more than 2 million words in 425 dictionaries in various fields: business, computer/Internet, medical, miscellaneous, religion, science, sports, technology, general, and slang.

In his e-mail of September 2, 1998, Robert Ware explained:

"On the personal side, I was almost entirely in contact with people who spoke one language and did not have much incentive to expand language abilities. Being in contact with the entire world has a way of changing that. And changing it for the better! [...] I have been slow to start including non-English dictionaries (partly because I am monolingual). But you will now find a few included."

A Web of Online Dictionaries, by Robert Beard, is an index of more than 800 on-line dictionaries in 150 languages, and other tools: multilingual dictionaries; specialized English dictionaries; thesauri and other vocabulary aids; language identifiers and guessers; an index of dictionary indices; a Web of on-line grammars; and a Web of linguistic fun (materials about linguistics for non-specialists).

Robert Beard answered my questions in his e-mail of September 1, 1998.

ML: "How do you see multilingualism on the Web?"

RB: "There was an initial fear that the Web posed a threat to multilingualism on the Web, since HTML and other programming languages are based on English and since there are simply more websites in English than any other language. However, my websites indicate that multilingualism is very much alive and the Web may, in fact, serve as a vehicle for preserving many endangered languages. I now have links to dictionaries in 150 languages and grammars of 65 languages. Moreover, the new attention paid by browser developers to the different languages of the world will encourage even more websites in different languages."

ML: "What did the use of the Internet bring to your professional life?"

RB: "As a language teacher, the Web represents a plethora of new resources produced by the target culture, new tools for delivering lessons (interactive Java and Shockwave exercises) and testing, which are available to students any time they have the time or interest — 24 hours a day, 7 days a week. It is also an almost limitless publication outlet for my colleagues and I, not to mention my institution."

ML: "How do you see the future of Internet-related activities as regards languages?"

RB: "Ultimately all course materials, including lecture notes, exercises, moot and credit testing, grading, and interactive exercises far more effective in conveying concepts that we have not even dreamed of yet. The Web will be an encyclopedia of the world by the world for the world. There will be no information or knowledge that anyone needs that will not be available. The major hindrance to international and interpersonal understanding, personal and institutional enhancement, will be removed. It would take a wilder imagination than mine to predict the effect of this development on the nature of humankind."

Initiated by the WorldWide Language Institute, NetGlos (The Multilingual Glossary of Internet Terminology) is currently being compiled from 1995 as a voluntary, collaborative project by a number of translators and other professionals. Versions for the following languages are being prepared: Chinese, Croatian, English, Dutch/Flemish, French, German, Greek, Hebrew, Italian, Maori, Norwegian, Portuguese, and Spanish.

Brian King, director of the WorldWide Language Institute, answered my questions in his e-mail of September 15, 1998.

ML: "How do you see multilingualism on the Web?"

BL: "Although English is still the most important language used on the Web, and the Internet in general, I believe that multilingualism is an inevitable part of the future direction of cyberspace.

Here are some of the important developments that I see as making a multilingual Web become a reality:

a) Popularization of information technology

Computer technology has traditionally been the sole domain of a 'techie' elite, fluent in both complex programming languages and in English — the universal language of science and technology. Computers were never designed to handle writing systems that couldn't be translated into ASCII. There wasn't much room for anything other than the 26 letters of the English alphabet in a coding system that originally couldn't even recognize acute accents and umlauts — not to mention nonalphabetic systems like Chinese.

But tradition has been turned upside down. Technology has been popularized. GUIs (graphical user interfaces) like Windows and Macintosh have hastened the process (and indeed it's no secret that it was Microsoft's marketing strategy to use their operating system to make computers easy to use for the average person). These days this ease of use has spread beyond the PC to the virtual, networked space of the Internet, so that now nonprogrammers can even insert Java applets into their webpages without understanding a single line of code.

b) Competition for a chunk of the 'global market' by major industry players

An extension of (local) popularization is the export of information technology around the world. Popularization has now occurred on a global scale and English is no longer necessarily the lingua franca of the user. Perhaps there is no true lingua franca, but only the individual languages of the users. One thing is certain — it is no longer necessary to understand English to use a computer, nor it is necessary to have a degree in computer science.

A pull from non-English-speaking computer users and a push from technology companies competing

for global markets has made localization a fast growing area in software and hardware development. This development has not been as fast as it could have been. The first step was for ASCII to become Extended ASCII. This meant that computers could begin to start recognizing the accents and symbols used in variants of the English alphabet — mostly used by European languages. But only one language could be displayed on a page at a time.

c) Technological developments

The most recent development is Unicode. Although still evolving and only just being incorporated into the latest software, this new coding system translates each character into 16 bytes. Whereas 8 byte Extended ASCII could only handle a maximum of 256 characters, Unicode can handle over 65,000 unique characters and therefore potentially accommodate all of the world's writing systems on the computer.

So now the tools are more or less in place. They are still not perfect, but at last we can at least surf the Web in Chinese, Japanese, Korean, and numerous other languages that don't use the Western alphabet. As the Internet spreads to parts of the world where English is rarely used — such as China, for example, it is natural that Chinese, and not English, will be the preferred choice for interacting with it. For the majority of the users in China, their mother tongue will be the only choice.

There is a change-over period, of course. Much of the technical terminology on the Web is still not translated into other languages. And as we found with our Multilingual Glossary of Internet Terminology — known as NetGlos — the translation of these terms is not always a simple process. Before a new term becomes accepted as the 'correct' one, there is a period of instability where a number of competing candidates are used. Often an English loanword becomes the starting point — and in many cases the endpoint. But eventually a winner emerges that becomes codified into published technical dictionaries as well as the everyday interactions of the nontechnical user. The latest version of NetGlos is the Russian one and it should be available in a couple of weeks or so [end of September 1998]. It will no doubt be an excellent example of the ongoing, dynamic process of 'Russification' of Web terminology.

d) Linguistic democracy

Whereas 'mother-tongue education' was deemed a human right for every child in the world by a UNESCO report in the early '50s, 'mother-tongue surfing' may very well be the Information Age equivalent. If the Internet is to truly become the Global Network that it is promoted as being, then all users, regardless of language background, should have access to it. To keep the Internet as the preserve of those who, by historical accident, practical necessity, or political privilege, happen to know English, is unfair to those who don't.

e) Electronic commerce

Although a multilingual Web may be desirable on moral and ethical grounds, such high ideals are not enough to make it other than a reality on a small-scale. As well as the appropriate technology being available so that the non-English speaker can go, there is the impact of 'electronic commerce' as a major force that may make multilingualism the most natural path for cyberspace.

Sellers of products and services in the virtual global marketplace into which the Internet is developing must be prepared to deal with a virtual world that is just as multilingual as the physical world. If they want to be successful, they had better make sure they are speaking the languages of their customers!"

ML: "What did the Internet bring to the life of your organization?"

BK: "Our main service is providing language instruction via the Web. Our company is in the unique position of having come into existence BECAUSE of the Internet!"

ML: "How do you see the future of Internet-related activities as regards languages?"

BK: "As a company that derives its very existence from the importance attached to languages, I believe the future will be an exciting and challenging one. But it will be impossible to be complacent about our successes and accomplishments. Technology is already changing at a frenetic pace. Life-long learning is a strategy that we all must use if we are to stay ahead and be competitive. This is a difficult enough task in an English-speaking environment. If we add in the complexities of interacting in a multilingual/multicultural cyberspace, then the task becomes even more demanding. As well as competition, there is also the necessity for cooperation — perhaps more so than ever before."

The seeds of cooperation across the Internet have certainly already been sown. Our NetGlos Project

has depended on the goodwill of volunteer translators from Canada, U.S., Austria, Norway, Belgium, Israel, Portugal, Russia, Greece, Brazil, New Zealand and other countries. I think the hundreds of visitors we get coming to the NetGlos pages everyday is an excellent testimony to the success of these types of working relationships. I see the future depending even more on cooperative relationships — although not necessarily on a volunteer basis."

3.4. Textual Databases

Let us take the example of two textual databases relating to the French language — the French FRANTEXT and the US-French ARTFL Project.

The FRANTEXT textual database has been available on the Web through subscription since the beginning of 1995. It is prepared in France by the Institut national de la langue française (INaLF) (National Institute of the French Language), a section of the Centre national de la recherche scientifique (CNRS) (National Center for Scientific Research). This interactive database includes 180 million words resulting from the automatic processing of a collection of 3,500 texts in arts, techniques and sciences, representing five centuries of literature (16th-20th centuries).

At the beginning of 1998, 82 research centers and university libraries in Europe, Australia, Canada and Japan were subscribing to FRANTEXT, with 1,250 work stations connected to the database, and about 50 questioning sessions per day. The detailed results of the inquiry sent to FRANTEXT users in January 1998 are presented on the website by Arlette Attali.

In the future, Arlette Attali is thinking about "contributing to the development of the linguistic tools associated to the FRANTEXT database and getting teachers, researchers and students to know them." In her e-mail of June 11, 1998, she also explained the changes brought by the Internet in her professional life:

"As I was more specially assigned to the development of textual databases at the INaLF, I had to explore the websites giving access to electronic texts and test them. I became a 'textual tourist' with the good and bad sides of this activity. The tendency to go quickly from one link to another, and to skip through the information, was a permanent danger — it is necessary to target what you are looking for if you don't want to lose your time. The use of the Web totally changed my working methods — my investigations are not only bookish and within a narrow circle anymore, on the contrary they are expanding thanks to the electronic texts available on the Internet."

The ARTFL Project (ARTFL: American and French Research on the Treasury of the French Language) is a cooperative project established in 1981 by the Institut national de la langue française (INaLF) (National Institute of the French Language, based in France) and the Division of the Humanities of the University of Chicago. Its purpose is to be a research tool for scholars and students in all areas of French studies.

The origin of the project is a 1957 initiative of the French government to create a new dictionary of the French language, the Trésor de la Langue Française (Treasure of the French Language). In order to provide access to a large body of word samples, it was decided to transcribe an extensive selection of French texts for use with a computer. Twenty years later, a corpus totaling some 150 million words had been created, representing a broad range of written French — from novels and poetry to biology and mathematics — stretching from the 17th to the 20th centuries.

This corpus of French texts was an important resource not only for lexicographers, but also for many other types of humanists and social scientists engaged in French studies — on both sides of the Atlantic. The result of this realization was the ARTFL Project, as explained on its website:

"At present the corpus consists of nearly 2,000 texts, ranging from classic works of French literature to various kinds of non-fiction prose and technical writing. The eighteenth, nineteenth and twentieth centuries are about equally represented, with a smaller selection of seventeenth century texts as well as some medieval and Renaissance texts. We have also recently added a Provençal database that includes 38 texts in their original spellings. Genres include novels, verse, theater, journalism, essays, correspondence, and treatises. Subjects include literary criticism, biology, history, economics, and philosophy. In most cases standard scholarly editions were used in converting the text into machine-readable form, and the data contain page references to these editions."

One of the largest of its kind in the world, the ARTFL database permits both the rapid exploration of single texts, and the inter-textual research of a kind. ARTFL is now on the Web, and the system is available through the Internet to its subscribers. Access to the database is organized through a consortium of user institutions, in most cases universities and colleges which pay an annual

subscription fee.

The ARTFL Encyclopédie Project is currently developing an on-line version of Diderot and d'Alembert's Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers, including all 17 volumes of text and 11 volumes of plates from the first edition, that is to say about 18,000 pages of text and exactly 20,736,912 words.

Published under the direction of Diderot between 1751 and 1772, the Encyclopédie counted as contributors the most prominent philosophers of the time: Voltaire, Rousseau, d'Alembert, Marmontel, d'Holbach, Turgot, etc.

"These great minds (and some lesser ones) collaborated in the goal of assembling and disseminating in clear, accessible prose the fruits of accumulated knowledge and learning. Containing 72,000 articles written by more than 140 contributors, the Encyclopédie was a massive reference work for the arts and sciences, as well as a machine de guerre which served to propagate Enlightened ideas [...] The impact of the Encyclopédie was enormous, not only in its original edition, but also in multiple reprintings in smaller formats and in later adaptations. It was hailed, and also persecuted, as the sum of modern knowledge, as the monument to the progress of reason in the eighteenth century. Through its attempt to classify learning and to open all domains of human activity to its readers, the Encyclopédie gave expression to many of the most important intellectual and social developments of its time."

At present, while work continues on the fully navigational, full-text version, ARTFL is providing public access on its website to the Prototype Demonstration of Volume One. From Autumn 1998 a preliminary version is released for consultation by all ARTFL subscribers.

Mentioned on the ARTFL home page in the Reference Collection, other ARTFL projects are: the 1st (1694) and 5th (1798) editions of the Dictionnaire de L'Académie française; Jean Nicot's Trésor de la langue française (1606) Dictionary; Pierre Bayle's Dictionnaire historique et critique (1740 edition) (text of an image-only version); The Wordsmyth English Dictionary-Thesaurus; Roget's Thesaurus, 1911 edition; Webster's Revised Unabridged Dictionary; the French Bible by Louis Segond and parallel Bibles in German, Latin, and English, etc.

Created by Michael S. Hart in 1971, the Project Gutenberg was the first information provider on the Internet. It is now the oldest digital library on the Web, and the biggest considering the number of works (1,500) which has been digitalized for it, with 45 new titles per month. Michael Hart's purpose is to put on the Web as many literary texts as possible for free.

In his e-mail of August 23, 1998, Michael S. Hart explained:

"We consider e-text to be a new medium, with no real relationship to paper, other than presenting the same material, but I don't see how paper can possibly compete once people each find their own comfortable way to e-texts, especially in schools. [...] My own personal goal is to put 10,000 e-texts on the Net, and if I can get some major support, I would like to expand that to 1,000,000 and to also expand our potential audience for the average e-text from 1.x% of the world population to over 10%... thus changing our goal from giving away 1,000,000,000,000 e-texts to 1,000 time as many... a trillion and a quadrillion in US terminology."

Project Gutenberg is now developing its foreign collections, as announced in the Newsletter of October 1997. In the Newsletter of March 1998, Michael S. Hart mentioned that Project Gutenberg's volunteers were now working on e-texts in French, German, Portuguese and Spanish, and he was also hoping to get some e-texts in the following languages: Arabic, Chinese, Danish, Dutch, Esperanto, Greek, Hebrew, Hungarian, Italian, Japanese, Korean, Latin, Lithuanian, Polish, Romanian, Russian, Slovak, Slovene, and Valencian (Catalan).

3.5. Terminological Databases

The free consultation of terminological databases on the Web is much appreciated by language specialists. There are some terminological databases maintained by international organizations, such as Eurodicautom, maintained by the Translation Service of the European Commission; ILOTTERM, maintained by the International Labour Organization (ILO), the ITU Telecommunication Terminology Database (TERMITE), maintained by the International Telecommunication Union (ITU) and the WHO Terminology Information System (WHOTERM), maintained by the World Health Organization (WHO).

Eurodicautom is the multilingual terminological database of the Translation Service of the European

Commission. Initially developed to assist in-house translators, it is consulted today by an increasing number of European Union officials other than translators, as well as by language professionals throughout the world. Its huge, constantly updated, contents is drafted in twelve languages (Danish, Dutch, English, Finnish, French, German, Greek, Italian, Latin, Portuguese, Spanish, Swedish), and covers a broad spectrum of human knowledge, while the main core relates to European Union topics.

ILOTERM is the quadrilingual (English, French, German, Spanish) terminology database maintained by the Terminology and Reference Unit of the Official Documentation Branch (OFFDOC) of the International Labour Office (ILO), Geneva, Switzerland. Its primary purpose is to provide solutions, reflecting current usage, to terminological problems in the social and labor fields. Terms are entered in English with their French, Spanish and/or German equivalents. The database also includes records (in up to four languages) concerning the structure and programmes of the ILO, official names of international institutions, national bodies and employers' and workers' organizations, as well as titles of international meetings and instruments.

The ITU Telecommunication Terminology Database (TERMITE) is maintained by the Terminology, References and Computer Aids to Translation Section of the Conference Department of the International Telecommunication Union (ITU), Geneva, Switzerland. TERMITE (59,000 entries) is a quadrilingual (English, French, Spanish, Russian) terminological database which contains all the terms which appeared in ITU printed glossaries since 1980, as well as more recent entries relating to the different activities of the Union.

Maintained by the World Health Organization (WHO), Geneva, Switzerland, the WHO Terminology Information System (WHOTERM) includes: the WHO General Dictionary Index, giving access to an English glossary of terms, with the French and Spanish equivalents for each term; three glossaries in English: Health for All, Programme Development and Management, and Health Promotion; the WHO TermWatch, an awareness service of the Technical Terminology, which is a service reflecting the current WHO usage — but not necessarily terms officially approved by WHO — and a series of links to health-related terminology

4. TRANSLATION RESOURCES

[In this chapter:]

[4.1. Translation Services / 4.2. Machine Translation / 4.3. Computer-Assisted Translation]

4.1. Translation Services

Maintained by Vorontsoff, Wesseling & Partners, Amsterdam, the Netherlands, Aquarius is a directory of translators and interpreters including 6,100 translators, 800 translation companies, 91 specialized areas of expertise and 369 language combinations. This non-commercial project helps to locate and contact the best translators in the world directly, without intermediaries or agencies. Aquarius Database can be searched using location, language combination and specialization.

Founded by Bill Dunlap, Euro-Marketing Associates proposes Global Reach, a methodology for companies to expand their Internet presence into a more international framework. this includes translating a website into other languages, actively promoting it and using local banner advertising to increase local website traffic in all on-line countries. Bill Dunlap explains:

"Promoting your website is at least as important as creating it, if not more important. You should be prepared to spend at least as much time and money in promoting your website as you did in creating it in the first place. With the "Global Reach" program, you can have it promoted in countries where English is not spoken, and achieve a wider audience... and more sales. There are many good reasons for taking the on-line international market seriously. "Global Reach" is a means for you to extend your website to many countries, speak to on-line visitors in their own language and reach on-line markets there."

In his e-mail of December 11, 1998, he also explains what the use of the Internet brought in his professional life:

"Since 1981, when my professional life started, I've been involved with bringing American companies in Europe. This is very much an issue of language, since the products and their marketing have to be in the languages of Europe in order for them to be visible here. Since the Web became popular in 1995 or so, I've turned these activities to their on-line dimension, and have come to champion European e-commerce among my fellow American compatriotes. Most lately at Internet World in New York, I spoke

about European e-commerce and how to use a website to address the various markets in Europe."

4.2. Machine Translation

Machine translation (MT) is the automated process of translating from one natural language to another. MT analyzes the language text in the source language and automatically generates corresponding text in the target language.

Characterized by the absence of any human intervention during the translation process, machine translation (MT) is also called "fully automatic machine translation (FAMT)". It differs from "machine-aided human translation (MAHT)" or "computer-assisted translation (CAT)", which involves some interaction between the translator and the computer.

As SYSTRAN, a company specialized in translation software, explains on its website:

"Machine translation software translates one natural language into another natural language. MT takes into account the grammatical structure of each language and uses rules to transfer the grammatical structure of the source language (text to be translated) into the target language (translated text). MT cannot replace a human translator, nor is it intended to."

The European Association for Machine Translation (EAMT) gives the following definition:

"Machine translation (MT) is the application of computers to the task of translating texts from one natural language to another. One of the very earliest pursuits in computer science, MT has proved to be an elusive goal, but today a number of systems are available which produce output which, if not perfect, is of sufficient quality to be useful for certain specific applications, usually in the domain of technical documentation. In addition, translation software packages which are designed primarily to assist the human translator in the production of translations are enjoying increasing popularity within professional translation organizations."

Machine translation is the earliest type of natural language processing. Here are the explanations given by Globalink:

"From the very beginning, machine translation (MT) and natural language processing (NLP) have gone hand-in-hand with the evolution of modern computational technology. The development of the first general-purpose programmable computers during World War II was driven and accelerated by Allied cryptographic efforts to crack the German Enigma machine and other wartime codes. Following the war, the translation and analysis of natural language text provided a testbed for the newly emerging field of Information Theory.

During the 1950s, research on Automatic Translation (known today as Machine Translation, or 'MT') took form in the sense of literal translation, more commonly known as word-for-word translations, without the use of any linguistic rules.

The Russian project initiated at Georgetown University in the early 1950s represented the first systematic attempt to create a demonstrable machine translation system. Throughout the decade and into the 1960s, a number of similar university and government-funded research efforts took place in the United States and Europe. At the same time, rapid developments in the field of Theoretical Linguistics, culminating in the publication of Noam Chomsky's *Aspects of the Theory of Syntax* (1965), revolutionized the framework for the discussion and understanding of the phonology, morphology, syntax and semantics of human language.

In 1966, the U.S. government-issued ALPAC report offered a prematurely negative assessment of the value and prospects of practical machine translation systems, effectively putting an end to funding and experimentation in the field for the next decade. It was not until the late 1970s, with the growth of computing and language technology, that serious efforts began once again. This period of renewed interest also saw the development of the Transfer model of machine translation and the emergence of the first commercial MT systems.

While commercial ventures such as SYSTRAN and METAL began to demonstrate the viability, utility and demand for machine translation, these mainframe-bound systems also illustrated many of the problems in bringing MT products and services to market. High development cost, labor-intensive lexicography and linguistic implementation, slow progress in developing new language pairs, inaccessibility to the average user, and inability to scale easily to new platforms are all characteristics of these second-generation systems."

A number of companies are specialized in machine translation development, such as Lernout &

Hauspie, Globalink, Logos or SYSTRAN.

Based in Ieper (Belgium) and Burlington (Massachusetts, USA), Lernout & Hauspie (L&H) is an international leader in the development of advanced speech technology for various commercial applications and products. The company offers four core technologies - automatic speech recognition (ASR), text-to-speech (TTS), text-to-text and digital speech compression. Its ASR, TTS and digital speech compression technologies are licensed to main companies in the telecommunications, computers and multimedia, consumer electronics and automotive electronics industries. Its text-to-text (translation) services are provided to information technology (IT) companies and vertical and automation markets.

The Machine Translation Group of Lernout & Hauspie comprises enterprises that develop, produce, and market highly sophisticated machine translation systems: L&H Language Technology, AppTek, AllLogic, NeocorTech and Globalink. Each is an international leader in its particular segment.

Founded in 1990, Globalink is a major U.S. company in language translation software and services, which offers customized translation solutions built around a range of software products, on-line options and professional translation services. The company publishes language translation software products in Spanish, French, Portuguese, German, Italian and English, and finds solutions to translation problems faced by individuals and small businesses, to multinational corporations and governments (a stand-alone product that gives a fast, draft translation or a full system to manage professional document translations). Globalink explains its corporate information on its website as follows:

"With Globalink's translation applications, the computer uses three sets of data: the input text, the translation program and permanent knowledge sources (containing a dictionary of words and phrases of the source language), and information about the concepts evoked by the dictionary and rules for sentence development. These rules are in the form of linguistic rules for syntax and grammar, and some are algorithms governing verb conjugation, syntax adjustment, gender and number agreement and word re-ordering.

Once the user has selected the text and set the machine translation process in motion the program begins to match words of the input text with those stored in its dictionary. Once a match is found, the application brings up a complete record that includes information on possible meanings of the word and its contextual relationship to other words that occur in the same sentence. The time required for the translation depends on the length of the text. A three-page, 750-word document takes about three minutes to render a first draft translation."

Randy Hobler is a Marketing Consultant for Globalink. He is currently acting as the Product Marketing Manager for Globalink's suite of Internet based products and services. In his e-mail of 3 September 1998, he wrote:

"85% of the content of the Web in 1998 is in English and going down. This trend is driven not only by more websites and users in non-English-speaking countries, but by increasing localization of company and organization sites, and increasing use of machine translation to/from various languages to translate websites.

Because the Internet has no national boundaries, the organization of users is bounded by other criteria driven by the medium itself. In terms of multilingualism, you have virtual communities, for example, of what I call 'Language Nations'... all those people on the Internet wherever they may be, for whom a given language is their native language. Thus, the Spanish Language nation includes not only Spanish and Latin American users, but millions of Hispanic users in the US, as well as odd places like Spanish-speaking Morocco.

Language Transparency: We are rapidly reaching the point where highly accurate machine translation of text and speech will be so common as to be embedded in computer platforms, and even in chips in various ways. At that point, and as the growth of the Web slows, the accuracy of language translation hits 98% plus, and the saturation of language pairs has covered the vast majority of the market, language transparency (any-language-to-any-language communication) will be too limiting a vision for those selling this technology. The next development will be 'transcultural, transnational transparency', in which other aspects of human communication, commerce and transactions beyond language alone will come into play. For example, gesture has meaning, facial movement has meaning and this varies among societies. The thumb-index finger circle means 'OK' in the United States. In Argentina, it is an obscene gesture.

When the inevitable growth of multi-media, multi-lingual videoconferencing comes about, it will be necessary to 'visually edit' gestures on the fly. The MIT Media Lab [MIT: Massachusetts Institute of Technology], Microsoft and many others are working on computer recognition of facial expressions,

biometric access identification via the face, etc. It won't be any good for a U.S. business person to be making a great point in a Web-based multi-lingual video conference to an Argentinian, having his words translated into perfect Argentinian Spanish if he makes the 'O' gesture at the same time. Computers can intercept this kind of thing and edit them on the fly.

There are thousands of ways in which cultures and countries differ, and most of these are computerizable to change as one goes from one culture to the other. They include laws, customs, business practices, ethics, currency conversions, clothing size differences, metric versus English system differences, etc., etc. Enterprising companies will be capturing and programming these differences and selling products and services to help the peoples of the world communicate better. Once this kind of thing is widespread, it will truly contribute to international understanding."

Logos is an international company (US, Canada and Europe) specialized in machine translation for 25 years, which provides various translation tools, machine translation systems and supporting services.

SYSTRAN (an acronym for System Translation) is a company specialized in machine translation software. SYSTRAN's headquarters are located in Soisy-sous-Montmorency, France. Sales and marketing, along with most development, operate out of its subsidiary, in La Jolla, California. The SYSTRAN site gives an interesting overview of the company's history. One of the company's products is AltaVista Translation, an automatic translation service of English Web pages into French, German, Italian, Portuguese, or Spanish, and vice versa, and is available on the AltaVista site, the most frequently used search engine on the Web.

Based in Montreal, Canada, Alis Technologies is an international company specialized in the development and marketing of language handling solutions and services, particularly at language implementation in the IT industry. Alis Translation Solutions (ATS) offers a wide selection of applications and languages, and multiple tools and services for best possible translation quality. Language Technology Solutions (LTS) is devoted to commercializing advanced tools and services in the field of language engineering and information technology. The unilingual information systems are transformed into software that users can put to work in their own language (90 languages covered).

Another machine translation development is SPANAM and ENGSPAN, which are fully automatic machine translation systems developed and maintained by the computational linguists, translators, and systems programmer of the Pan American Health Organization (PAHO), Washington, D.C. The PAHO Translation Unit has used SPANAM (Spanish to English) and ENGSPAN (English to Spanish) to process over 25 million words since 1980. Staff and free-lance translators postedit the raw output to produce high-quality translations with a 30-50% gain in productivity. The system is installed on a local area network at PAHO Headquarters and is used regularly by staff in the technical and administrative units. The software is also installed in a number of PAHO field offices and has been licensed to public and non-profit institutions in the US, Latin America, and Spain.

Some associations also contribute to machine translation development.

The Association for Computational Linguistics (ACL) is the main international scientific and professional society for people working on problems involving natural language and computation. Published by MIT Press, the ACL quarterly journal, Computational Linguistics (ISSN 0891-2017), continues to be the primary forum for research on computational linguistics and natural language processing. The Finite String is its newsletter supplement. The European branch of ACL is the European Chapter of the Association of Computational Linguistics (EACL), which provides a regional focus for its members.

The International Association for Machine Translation (IAMT) heads a worldwide network with three regional components: the Association for Machine Translation in the Americas (AMTA), the European Association for Machine Translation (EAMT) and the Asia-Pacific Association for Machine Translation (AAMT).

The Association for Machine Translation in the Americas (AMTA) presents itself as an association dedicated to anyone interested in the translation of languages using computers in some way. It has members in Canada, Latin America, and the United States. This includes people with translation needs, commercial system developers, researchers, sponsors, and people studying, evaluating, and understanding the science of machine translation and educating the public on important scientific techniques and principles involved.

The European Association for Machine Translation (EAMT) is based in Geneva, Switzerland. This organization serves the growing community of people interested in MT (machine translation) and translation tools, including users, developers, and researchers of this increasingly viable technology.

The Asia-Pacific Association for Machine Translation (AAMT), formerly called the Japan Association for Machine Translation (created in 1991), is comprised of three entities: researchers, manufacturers, and users of machine translation systems. The association endeavors to develop machine translation technologies to expand the scope of effective global communications and, for this purpose, is engaged in machine translation system development, improvement, education, and publicity.

In Web embraces language translation, an article of ZDNN (ZD Network News) of July 21, 1998, Martha L. Stone explains:

"Among the new products in the \$10 billion language translation business are instant translators for websites, chat rooms, e-mail and corporate intranets.

The leading translation firms are mobilizing to seize the opportunities. Such as:

SYSTRAN has partnered with AltaVista and reports between 500,000 and 600,000 visitors a day on babelfish.altavista.digital.com, and about 1 million translations per day — ranging from recipes to complete Web pages.

About 15,000 sites link to babelfish, which can translate to and from French, Italian, German, Spanish and Portuguese. The site plans to add Japanese soon.

'The popularity is simple. With the Internet, now there is a way to use US content. All of these contribute to this increasing demand,' said Dimitros Sabatakakis, group CEO of SYSTRAN, speaking from his Paris home.

Alis technology powers the Los Angeles Times' soon-to-be launched language translation feature on its site. Translations will be available in Spanish and French, and eventually, Japanese. At the click of a mouse, an entire web page can be translated into the desired language.

Globalink offers a variety of software and Web translation possibilities, including a free e-mail service and software to enable text in chat rooms to be translated.

But while these so-called 'machine' translations are gaining worldwide popularity, company execs admit they're not for every situation.

Representatives from Globalink, Alis and SYSTRAN use such phrases as 'not perfect' and 'approximate' when describing the quality of translations, with the caveat that sentences submitted for translation should be simple, grammatically accurate and idiom-free.

'The progress on machine translation is moving at Moore's Law — every 18 months it's twice as good,' said Vin Crosbie, a Web industry analyst in Greenwich, Conn. 'It's not perfect, but some [non-English speaking] people don't realize I'm using translation software.'

With these translations, syntax and word usage suffer, because dictionary-driven databases can't decipher between homonyms — for example, 'light' (as in the sun or light bulb) and 'light' (the opposite of heavy).

Still, human translation would cost between \$50 and \$60 per Web page, or about 20 cents per word, SYSTRAN's Sabatakakis said.

While this may be appropriate for static 'corporate information' pages, the machine translations are free on the Web, and often less than \$100 for software, depending on the number of translated languages and special features."

4.3. Computer-Assisted Translation

Within the World Health Organization (WHO), Geneva, Switzerland, the Computer-assisted Translation and Terminology (Unit (CTT) is assessing technical options for using computer-assisted translation (CAT) systems based on "translation memory". With such systems, translators have immediate access to previous translations of portions of the text before them. These reminders of previous translations can be accepted, rejected or modified, and the final choice is added to the memory, thus enriching it for future reference. By archiving daily output, the translator would soon have access to an enormous "memory" of ready-made solutions for a considerable number of translation problems. Several projects are currently under way in such areas as electronic document archiving and retrieval, bilingual/multilingual text alignment, computer-assisted translation, translation memory and terminology database management, and speech recognition.

Contrary to the imminent outbreak of the universal translation machine announced some 50 years ago, the machine translation systems don't yet produce good quality translations. Why not? Pierre Isabelle and Patrick Andries, from the Laboratoire de recherche appliquée en linguistique informatique (RALI) (Laboratory for Applied Research in Computational Linguistics) in Montreal, Quebec, explain this failure in *La traduction automatique, 50 ans après* (Machine translation, 50 years later), an article published in the *Dossiers* of the daily cybermagazine *Multimédium*:

"The ultimate goal of building a machine capable of competing with a human translator remains elusive due to the slow progress of the research. [...] Recent research, based on large collections of texts called corpora - using either statistical or analogical methods - promise to reduce the quantity of manual work required to build a MT [machine translation] system, but it is less sure than they can promise a substantial improvement in the quality of machine translation. [...] the use of MT will be more or less restricted to information assimilation tasks or tasks of distribution of texts belonging to restricted sub-languages."

According to Yehochua Bar-Hillel's ideas expressed in *The State of Machine Translation*, an article published in 1951, Pierre Isabelle and Patrick Andries define three MT implementation strategies: 1) a tool of information assimilation to scan multilingual information and supply rough translation, 2) situations of "restricted language" such as the METEO system which, since 1977, has been translating the weather forecasts of the Canadian Ministry of Environment, 3) the human being/machine coupling before, during and after the MT process, which is not inevitably economical compared to traditional translation.

The authors favour "a workstation for the human translator" more than a "robot translator":

"The recent research on the probabilist methods permitted in fact to demonstrate that it was possible to modelize in a very efficient way some simple aspects of the translation relationship between two texts. For example, methods were set up to calculate the correct alignment between the text sentences and their translation, that is, to identify the sentence(s) of the source text which correspond(s) to each sentence of the translation. Applied on a large scale, these techniques allow the use of archives of a translation service to build a translation memory which will often permit the recycling of previous translation fragments. Such systems are already available on the translation market (IBM Translation Manager II, Trados Translator's Workbench by Trados, RALI TransSearch, etc.)

The most recent research focuses on models able to automatically set up the correspondences at a finer level than the sentence level: syntagms and words. The results obtained foresee a whole family of new tools for the human translator, including aids for terminological studying, aids for dictation and translation typing, and detectors of translation errors."

5. LANGUAGE-RELATED RESEARCH

[In this chapter:]

[5.1. Machine Translation Research / 5.2. Computational Linguistics / 5.3. Language Engineering / 5.4. Internationalization and Localization]

5.1. Machine Translation Research

The CL/MT Research Group (Computational Linguistics (CL) and Machine Translation (MT) Group) is a research group in the Department of Language and Linguistics at the University of Essex, United Kingdom. It serves as a focus for research in computational, and computationally oriented, linguistics. It has been in existence since the late 1980s, and has played a role in a number of important computational linguistics research projects.

Founded in 1986, the Center for Machine Translation (CMT) is now a research center within the new Language Technologies Institute at the School of Computer Science at Carnegie Mellon University (CMU), Pittsburgh, Pennsylvania. It conducts advanced research and development in a suite of technologies for natural language processing, with a primary focus on high-quality multilingual machine translation.

Within the CLIPS Laboratory (CLIPS: Communication langagière et interaction personne-système = Language Communication and Person-System Communication) of the French IMAG Federation, the Groupe d'étude pour la traduction automatique (GETA) (Study Group for Machine Translation) is a multi-disciplinary team of computer scientists and linguists. Its research topics concern all the theoretical, methodological and practical aspects of computer-assisted translation (CAT), or more

generally of multilingual computing. The GETA participates in the UNL (Universal Networking Language) project, initiated by the Institute of Advanced Studies (IAS) of the United Nations University (UNU).

"UNL (Universal Networking Language) is a language that - with its companion "enconverter" and "deconverter" software - enables communication among peoples of differing native languages. It will reside, as a plug-in for popular World Wide Web browsers, on the Internet, and will be compatible with standard network servers. The technology will be shared among the member states of the United Nations. Any person with access to the Internet will be able to "enconvert" text from any native language of a member state into UNL. Just as easily, any UNL text can be "deconverted" from UNL into native languages. United Nations University's UNL Center will work with its partners to create and promote the UNL software, which will be compatible with popular network servers and computing platforms."

The Natural Language Group (NLG) at the Information Sciences Institute (ISI) of the University of Southern California (USC) is currently involved in various aspects of computational/natural language processing. The group's projects are: machine translation; automated text summarization; multilingual verb access and text management; development of large concept taxonomies (ontologies); discourse and text generation; construction of large lexicons for various languages; and multimedia communication.

Eduard Hovy, Head of the Natural Language Group, explained in his e-mail of August 27, 1998:

"Your presentation outline looks very interesting to me. I do wonder, however, where you discuss the language-related applications/functionalities that are not translation, such as information retrieval (IR) and automated text summarization (SUM). You would not be able to find anything on the Web without IR! — all the search engines (AltaVista, Yahoo!, etc.) are built upon IR technology. Similarly, though much newer, it is likely that many people will soon be using automated summarizers to condense (or at least, to extract the major contents of) single (long) documents or lots of (any length) ones together. [...]

In this context, multilingualism on the Web is another complexifying factor. People will write their own language for several reasons — convenience, secrecy, and local applicability — but that does not mean that other people are not interested in reading what they have to say! This is especially true for companies involved in technology watch (say, a computer company that wants to know, daily, all the Japanese newspaper and other articles that pertain to what they make) or some Government Intelligence agencies (the people who provide the most up-to-date information for use by your government officials in making policy, etc.). One of the main problems faced by these kinds of people is the flood of information, so they tend to hire 'weak' bilinguals who can rapidly scan incoming text and throw out what is not relevant, giving the relevant stuff to professional translators. Obviously, a combination of SUM and MT (machine translation) will help here; since MT is slow, it helps if you can do SUM in the foreign language, and then just do a quick and dirty MT on the result, allowing either a human or an automated IR-based text classifier to decide whether to keep or reject the article.

For these kinds of reasons, the US Government has over the past five years been funding research in MT, SUM, and IR, and is interested in starting a new program of research in Multilingual IR. This way you will be able to one day open Netscape or Explorer or the like, type in your query in (say) English, and have the engine return texts in *all* the languages of the world. You will have them clustered by subarea, summarized by cluster, and the foreign summaries translated, all the kinds of things that you would like to have.

You can see a demo of our version of this capability, using English as the user language and a collection of approx. 5,000 texts of English, Japanese, Arabic, Spanish, and Indonesian, by visiting MuST Multilingual Information Retrieval, Summarization, and Translation System.

Type your query word (say, 'baby', or whatever you wish) in and press 'Enter/Return'. In the middle window you will see the headlines (or just keywords, translated) of the retrieved documents. On the left you will see what language they are in: 'Sp' for Spanish, 'Id' for Indonesian, etc. Click on the number at left of each line to see the document in the bottom window. Click on 'Summarize' to get a summary. Click on 'Translate' for a translation (but beware: Arabic and Japanese are extremely slow! Try Indonesian for a quick word-by-word 'translation' instead).

This is not a product (yet); we have lots of research to do in order to improve the quality of each step. But it shows you the kind of direction we are heading in."

"How do you see the future of Internet-related activities as regards languages?"

"The Internet is, as I see it, a fantastic gift to humanity. It is, as one of my graduate students recently said, the next step in the evolution of information access. A long time ago, information was transmitted orally only; you had to be face-to-face with the speaker. With the invention of writing, the time barrier broke down — you can still read Seneca and Moses. With the invention of the printing press, the access barrier was overcome — now *anyone* with money to buy a book can read Seneca and Moses. And today, information access becomes almost instantaneous, globally; you can read Seneca and Moses from your computer, without even knowing who they are or how to find out what they wrote; simply open AltaVista and search for 'Seneca'. This is a phenomenal leap in the development of connections between people and cultures. Look how today's Internet kids are incorporating the Web in their lives.

The next step? — I imagine it will be a combination of computer and cellular phone, allowing you as an individual to be connected to the Web wherever you are. All your diary, phone lists, grocery lists, homework, current reading, bills, communications, etc., plus AltaVista and the others, all accessible (by voice and small screen) via a small thing carried in your purse or on your belt. That means that the barrier between personal information (your phone lists and diary) and non-personal information (Seneca and Moses) will be overcome, so that you can get to both types anytime. I would love to have something that tells me, when next I am at a conference and someone steps up, smiling to say hello, who this person is, where last I met him/her, and what we said then!

But that is the future. Today, the Web has made big changes in the way I shop (I spent 20 minutes looking for plane routes for my next trip with a difficult transition on the Web, instead of waiting for my secretary to ask the travel agent, which takes a day). I look for information on anything I want to know about, instead of having to make a trip to the library and look through complicated indexes. I send e-mail to you about this question, at a time that is convenient for me, rather than your having to make a phone appointment and then us talking for 15 minutes. And so on."

The Computing Research Laboratory (CRL) at New Mexico State University (NMSU) is a non-profit research enterprise committed to basic research and software development in advanced computing applications concentrated in the areas of natural language processing, artificial intelligence and graphical user interface design. Applications developed from basic research endeavors include a variety of configurations of machine translation, information extraction, knowledge acquisition, intelligent teaching, and translator workstation systems.

Maintained by the Department of Linguistics of the Translation Research Group of Brigham Young University (BYU), Utah, TTT.org (Translation, Theory and Technology) provides information about language theory and technology, particularly relating to translation. Translation technology includes translator workbench tools and machine translation. In addition to translation tools, TTT.org is interested in data exchange standards that allow various tools to interoperate, allowing the integration of tools from multiple vendors in the multilingual document production chain.

In the area of data exchange standards, TTT.org is actively involved in the development of MARTIF (machine-readable terminology interchange format). MARTIF is a format to facilitate the interchange of terminological data among terminology management systems. This format is the result of several years of intense international collaboration among terminologists and database experts from various organizations, including academic institutions, the Text Encoding Initiative (TEI), and the Localisation Industry Standards Association (LISA).

5.2. Computational Linguistics

The Laboratoire de recherche appliquée en linguistique informatique (RALI) (Laboratory of Applied Research in Computational Linguistics) is a laboratory of the University of Montreal, Quebec. The RALI's personnel includes experienced computer scientists and linguists in natural language processing both in classical symbolic methods as well as in newer probabilist methods.

Thanks to the Incognito laboratory, which was founded in 1983, the University of Montreal's Computer Science and Operational Research Department (DIRO) established itself as a leading research centre in the area of natural language processing. In June 1997, Industry Canada agreed to transfer to the DIRO all the activities of the machine-aided translation program (TAO), which had been conducted at the Centre for Information Technology Innovation (CITI) since 1984. A new laboratory — the RALI — was opened in order to promote and develop the results of the CITI's research, allowing the members of the former TAO team to pursue their work within the university community. The RALI's areas of expertise include work in: automatic text alignment, automatic text generation, automatic reaccentuation, language identification and finite state transducers.

The RALI produces the "TransX family" of what it calls "a new generation" of translation support tools (TransType, TransTalk, TransCheck and TransSearch), which are based on probabilistic translation models that automatically calculate the correspondences between the text produced by a translator and the original source language text.

" TransType speeds up the keying-in of a translation by anticipating a translator's choices and critiquizing them when appropriate. In proposing its suggestions, TransType takes into account both the source text and the partial translation that the translator has already produced.

TransTalk is an automatic dictation system that makes use of a probabilistic translation model in order to improve the performance of its voice recognition model.

TransCheck automatically detects certain types of translation errors by verifying that the correspondences between the segments of a draft and the segments of the source text respect well-known properties of a good translation.

TransSearch allows translators to search databases of pre-existing translations in order to find ready-made solutions to all sorts of translation problems. In order to produce the required databases, the translations and the source language texts must first be aligned."

Some of RALI's other projects are:

- the SILC Project, concerning language identification. When a document is submitted to the system, SILC attempts to determine what language the document is written in and the character set in which it is encoded.

- the FAP: Finite Automata Package (FAP), a project concerning finite-state transducers. The finite-state automaton is a simple and efficient computational device for describing sequences of symbols (words, characters, etc.) known as the regular languages. The finite-state transducer is a device for linking pairs of these sequences under the control of a grammar of local correspondences, and thus provides a means of rewriting one sequence as another. Applications of these techniques in NLP include: dictionaries, morphological analysis, part-of-speech tagging, syntactic analysis, and speech processing.

The Xerox Palo Alto Research Center (PARC)'s projects include two main projects concerning languages: Inter-Language Unification (ILU) and Natural Language Theory and Technology (NLTT).

The Inter-Language Unification (ILU) System is a multi-language object interface system. The object interfaces provided by ILU hide implementation distinctions between different languages, between different address spaces, and between operating system types. ILU can be used to build multilingual object-oriented libraries ("class libraries") with well-specified language-independent interfaces. It can also be used to implement distributed systems, or to define and document interfaces between the modules of non-distributed programs.

The goal of Natural Language Theory and Technology (NLTT) is to develop theories of how information is encoded in natural language and technologies for mapping information to and from natural language representations. This will enable the efficient and intelligent handling of natural language text in critical phases of document processing, such as recognition, summarizing, indexing, fact extraction and presentation, document storage and retrieval, and translation. It will also increase the power and convenience of communicating with machines in natural language.

Based in Cambridge, United Kingdom, and Grenoble, France, The Xerox Research Centre Europe (XRCE) is also a research organization of the international company XEROX, which focuses on increasing productivity in the workplace through new document technologies, with several tools and projects relating to languages.

One of Xerox's research activities is MultiLingual Theory and Technology (MLTT), to study how to analyze and generate text in many languages (English, French, German, Italian, Spanish, Russian, Arabic, etc.). The MLTT team creates basic tools for linguistic analysis, e.g. morphological analysers, parsing and generation platforms and corpus analysis tools. These tools are used to develop descriptions of various languages and the relation between them. Currently under development are phrasal parsers for French and German, a lexical functional grammar (LFG) for French and projects on multilingual information retrieval, translation and generation.

Founded in 1979, the American Association for Artificial Intelligence (AAAI) is a non-profit scientific society devoted to advancing the scientific understanding of the mechanisms underlying thought and intelligent behavior and their embodiment in machines. AAAI also aims to increase public understanding of artificial intelligence, improve the teaching and training of AI practitioners, and

provide guidance for research planners and funders concerning the importance and potential of current AI developments and future directions.

The Institut Dalle Molle pour les études sémantiques et cognitives (ISSCO) (Dalle Molle Institute for Semantic and Cognitive Studies) is a research laboratory attached to the University of Geneva, Switzerland, which conducts basic and applied research in computational linguistics (CL), and artificial intelligence (AI). The site gives a presentation of the ISSCO projects (European projects, projects of the Swiss National Science Foundation, projects of the French-speaking community, etc.).

Created by the Foundation Dalle Molle in 1972 for research into cognition and semantics, ISSCO has come to specialize in natural language processing and, in particular, in multilingual language processing, in a number of areas : machine translation, linguistic environments, multilingual generation, discourse processing, data collection, etc. The University of Geneva provides administrative support and infrastructure for ISSCO. The research is funded solely by grants and by contracts with public and private bodies.

ISSCO is multi-disciplinary and multi-national, "drawing its staff and its visitors from the disciplines of computer science, linguistics, mathematics, psychology and philosophy. The long-term staff of the Institute is relatively small in number; with a much larger number of visitors coming for stays ranging from a month to two years. This ensures a continual exchange of ideas and encourages flexibility of approach amongst those associated with the Institute."

The International Conferences on Computational Linguistics (COLINGs) are organized every two years by the International Committee on Computational Linguistics (ICCL).

"The International Committee on Computational Linguistics was set up by David Hays in the mid-Sixties as a permanent body to run international computational linguistics conferences in an original way, with no permanent secretariat, subscriptions or funds. It was ahead of its time in that and other ways. COLING has always been distinguished by pleasant venues and atmosphere, rather than by the clinical efficiency of an airport conference hotel: COLINGs are simply nice conferences to be at. [...] In recent years, the ACL [Association for Computational Linguistics] has given great assistance and cooperation in keeping COLING proceedings available and distributed."

5.3. Language Engineering

Launched in January 1999 by the European Commission, the website HLTCentral (HLT: Human Language Technologies) gives a short definition of language engineering:

"Through language engineering we can find ways of living comfortably with technology. Our knowledge of language can be used to develop systems that recognise speech and writing, understand text well enough to select information, translate between different languages, and generate speech as well as the printed world.

By applying such technologies we have the ability to extend the current limits of our use of language. Language enabled products will become an essential and integral part of everyday life."

A full presentation of language engineering can be found in Language Engineering: Harnessing the Power of Language.

From 1992 to 1998, the Language Engineering Sector was part of the Telematics Applications Programme of the European Commission. Its aim was to facilitate the use of telematics applications and to increase the possibilities for communication in and between European languages. RTD (research and technological development) work focused on pilot projects that integrated language technologies into information and communications applications and services. A key objective was to improve their ease of use and functionality and broaden their scope across different languages.

From January 1999, the Language Engineering Sector has been rebranded as Human Language Technologies (HLT), a sector of the IST Programme (IST: Information Society Technologies) of the European Commission for 1999-2002. HLTCentral has been set up by the LINGLINK Project as the springboard for access to Language Technology resources on the Web: information, news, downloads, links, events, discussion groups and a number of specially-commissioned studies (e-commerce, telecommunications, Call Centres, Localization, etc.).

The Multilingual Application Interface for Telematic Services (MAITS) is a consortium formed to specify an applications programming interface (API) for multilingual applications in the telematic services. A number of telematic applications, such as X.500, WWW, X.400, internet mail and data bases,

is planned to be enhanced to use this i18n API, and products are planned to be implemented using the API.

FRANCIL (Réseau francophone de l'ingénierie de la langue) (Francophone Network in Language Engineering) is a programme launched in June 1994 by the Agence universitaire de la francophonie (AUPELF-UREF) (University Agency for Francophony) to strengthen activities in linguistic engineering, particularly for automatic language processing. This quickly-growing sector includes research and development for text analysis and generation, and for speech recognition, comprehension and synthesis. It also includes some applications in the following fields: document management, communication between the human being and the machine, writing aid, and computer-assisted translation.

5.4. Internationalization and Localization

"Towards communicating on the Internet in any language..." Babel is an Alis Technologies/ Internet Society joint project to internationalize the Internet. Its multilingual site (English, French, German, Italian, Portuguese, Spanish and Swedish) has two main sections: languages (the world's languages; typographical and linguistic glossary; Francophonie (French-speaking countries); and the Internet and multilingualism (developing your multilingual Web site; coding the world's writing).

The Localisation Industry Standards Association (LISA) is a main organization for the localization and internationalization industry. The current membership of 130 leading players from all around the world includes software publishers, hardware manufacturers, localization service vendors, and an increasing number of companies from related IT sectors. LISA defines its mission as "promoting the localization and internationalization industry and providing a mechanism and services to enable companies to exchange and share information on the development of processes, tools, technologies and business models connected with localization, internationalization and related topics". Its site is housed and maintained by the University of Geneva, Switzerland.

W3C Internationalization/Localization is part of the World Wide Web Consortium (W3C), an international industry consortium founded in 1994 to develop common protocols for the World Wide Web. The site gives in particular a definition of protocols used for internationalization/localization: HTML; base character set; new tags and attributes; HTTP; language negotiation; URLs & other identifiers including non-ASCII characters; etc. It also offers some help with creating a multilingual site.

6. INDEX OF WEBSITES

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 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)

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

Computer-Assisted Translation and Terminology Unit (CTT) / World Health Organization (WHO)

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 / Swiss Federal Administration

ENGSPAN (SPANAM and ENGSPAN) / Pan American Health Organisation (PAHO)

Ethnologue (The)

Eurodicautom / European Commission

EUROCALL (European Association for Computer-Assisted Language Learning)

European Association for Machine Translation (EAMT)

European Chapter of the Association of Computational Linguistics (EACL)

European Committee for the Respect of Cultures and Languages in Europe (ECRCLE)

European Language Resources Association (ELRA)

European Minority Languages / Sabhal Mór Ostaig

European Network in Language and Speech (ELNET)

Fonds francophone des inforoutes / Agence de la francophonie

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 Technologies (HLTCentral) / European Commission

Human-Languages Page (The)

ILOTTERM / International Labour Organization (ILO)

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) (Logos Home Page)

Lernout & Hauspie

LINGUIST List (The)

Localisation Industry Standards Association (LISA)

Logos (Canada, USA, Europe)

Logos (Italy)

Logos Home Page (The Languages of the World by Computers and the Internet)

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) / European Commission

MultiLingual Theory and Technology (MLTT) / Xerox Research Centre Europe (XRCE)

Multilingual Tools and Services / European Union

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

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 / Pan American Health Organization (PAHO)

Speech on the Web

SYSTRAN

TERMITE (ITU Telecommunication Terminology Database) / International Telecommunication Union (ITU)

Travlang

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

Universal Networking Language (UNL) / United Nations University (UNU)

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

Web Languages Hit Parade / Babel

Web of Online Dictionaries (A)

WELL (Web Enhanced Language Learning)

WHOTERM (WHO Terminology Information System) / World Health Organization (WHO)

Xerox Palo Alto Research Center (PARC)

Xerox Research Centre Europe (XRCE)

Yamada WWW Language Guides

7. INDEX OF NAMES

An asterisk (*) indicates the persons who sent contributions especially for this study.

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 (Ministry of Culture and Communications in Quebec)

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

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* (The 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 Computers and the Internet)

Caoimhín P. Ó Donnai* (European Minority Languages)

Henri Slettenhaar* (professor at the 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, Italy)

Robert Ware* (2) (OneLook Dictionaries)

Copyright © 1999 Marie Lebert

*** END OF THE PROJECT GUTENBERG EBOOK MULTILINGUALISM ON THE 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.

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. 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), 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.

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

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.

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

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.

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.