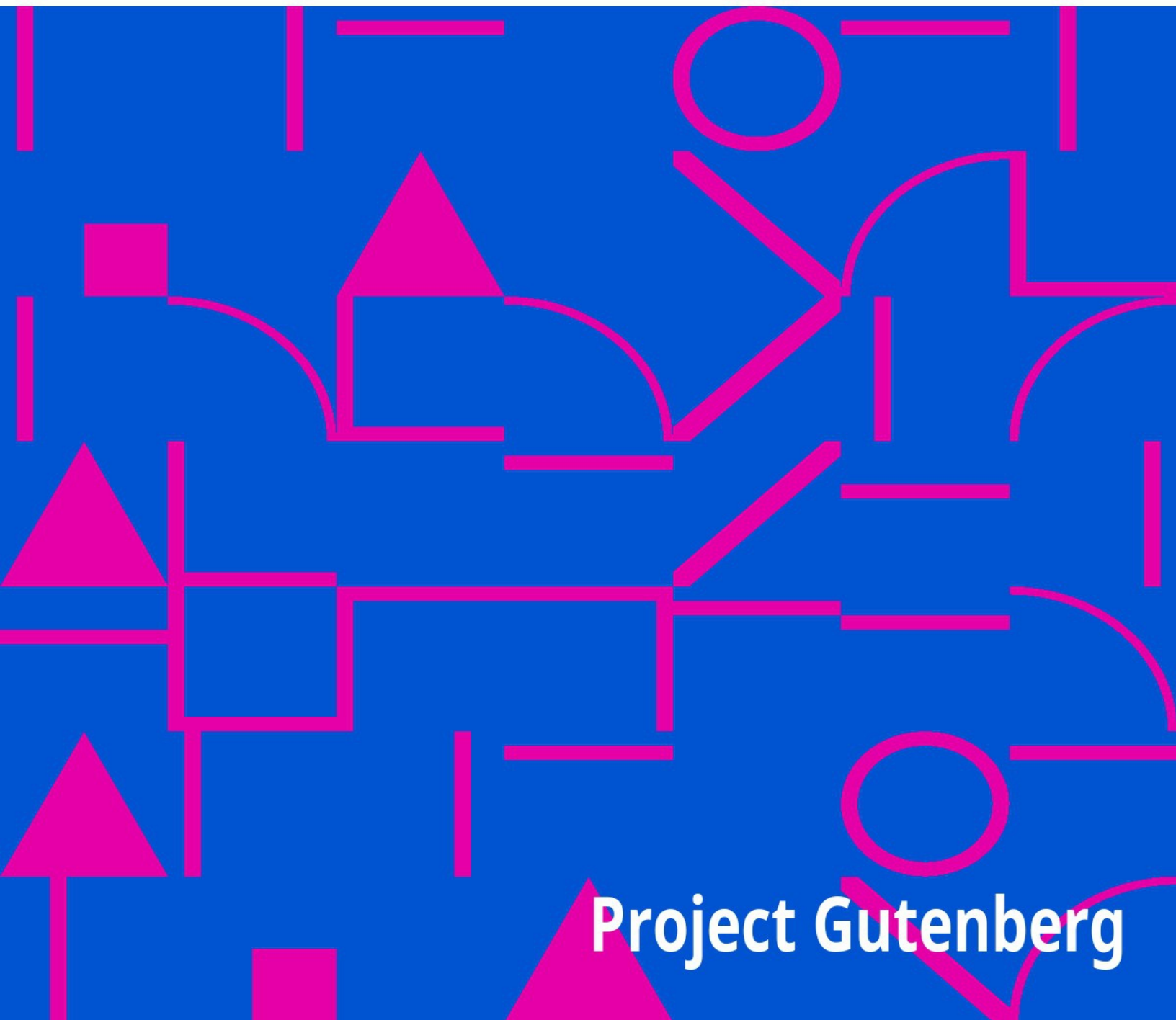


The web, a multilingual encyclopedia

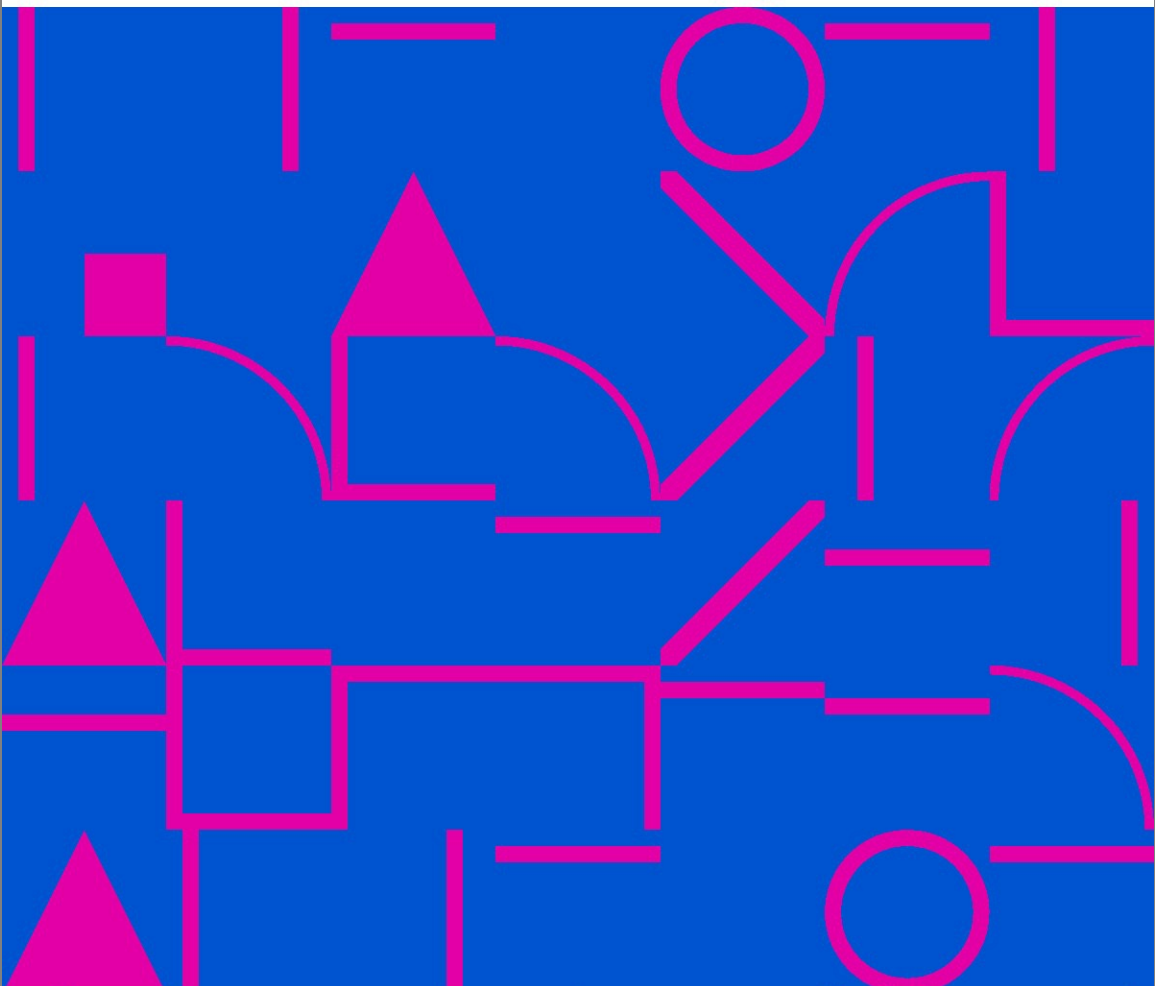
Marie Lebert

An abstract geometric pattern on a blue background, featuring various pink shapes including lines, squares, triangles, circles, and arcs, some of which are interconnected to form a network-like structure.

Project Gutenberg

The web, a multilingual encyclopedia

Marie Lebert



The logo for Project Gutenberg is a horizontal banner. It has a blue background with a pink geometric pattern consisting of a vertical line, a square, a triangle, and a circle. The text "Project Gutenberg" is written in white, sans-serif font across the center of the banner.

Project Gutenberg

The Project Gutenberg eBook of The web, a multilingual encyclopedia

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.

*** This is a COPYRIGHTED Project Gutenberg eBook. Details Below. ***

*** Please follow the copyright guidelines in this file. ***

Title: The web, a multilingual encyclopedia

Author: Marie Lebert

Release date: January 10, 2013 [eBook #41814]

Language: English

*** START OF THE PROJECT GUTENBERG EBOOK THE WEB, A
MULTILINGUAL ENCYCLOPEDIA ***

Produced by Al Haines

THE WEB, A MULTILINGUAL ENCYCLOPEDIA

MARIE LEBERT, 2012

TABLE OF CONTENTS

1974 > The internet “took off” 1990 > The invention of the web 1990 > The LINGUIST List 1991 > From ASCII to Unicode 1994 > Travlang, travel and languages 1995 > The Internet Dictionary Project 1995 > NetGlos, a glossary of the internet 1995 > Various languages on our screen 1995 > Global Reach, promoting localization 1996 > OneLook Dictionaries, a “fast finder” 1997 > 82.3% of the web in English 1997 > The internet, a tool for minority languages 1997 > A European terminology database 1997 > Babel Fish, a free translation software 1997 > The tools of the translation company Logos 1997 > Specialized terminology databases 1998 > The need for a “linguistic democracy” 1999

> Bilingual dictionaries in WordReference.com 1999 > The internet, a mandatory tool for translators 1999 > The need for bilingual information online 2000 > Online encyclopedias and dictionaries 2000 > The web portal yourDictionary.com 2000 > Project Gutengerg and languages 2001 > Wikipedia, a collaborative encyclopedia 2001 > UNL, a digital metalanguage project 2001 > A market for language translation software 2004 > The web 2.0, community and sharing 2007 > The ISO 639-3 standard to identify languages 2007 > Google Translate 2009 > 6,909 languages in the Ethnologue 2010 > A UNESCO atlas for endangered languages

INTRODUCTION

"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." (Robert Beard, founder of A Web of Online Dictionaries, september 1998)

This book is a chronology in 31 chapters from 1974 to 2010. Many thanks to all those who are quoted here, for their time and their friendship. Unless specified otherwise, the quotes are excerpts from the interviews conducted by the author during several years and published in the same collection.

1974 > THE INTERNET "TOOK OFF"

[Summary] The internet “took off” in 1974 with the creation of TCP/IP (Transmission Control Protocol / Internet Protocol) by Vinton Cerf and Bob Kahn, fifteen years before the invention of the web. The internet expanded as a network linking U.S. governmental agencies, universities and research centers, before spreading worldwide in 1983. The internet got its first boost in 1990 with the invention of the web by Tim Berners-Lee, and its second

boost in 1993 with the release of Mosaic, the first browser for the general public. The Internet Society (ISOC) was founded in 1992 by Vinton Cerf to promote the development of the internet as a medium that was becoming part of our lives. There were 100 million internet users in December 1997, with one million new users per month, and 300 million users in December 2000.

The internet “took off” in 1974 with the creation of TCP/IP (Transmission Control Protocol / Internet Protocol) by Vinton Cerf and Bob Kahn, fifteen years before the invention of the web.

A new medium

The internet expanded as a network linking U.S. governmental agencies, universities and research centers, before spreading worldwide in 1983.

The internet got its first boost in 1990 with the invention of the web by Tim Berners-Lee, and its second boost in 1993 with the release of Mosaic, the first browser for the general public.

Vinton Cerf founded the Internet Society (ISOC) in 1992 to promote the development of the internet as a medium that was becoming part of our lives. When interviewed by the French daily Libération on 16 January 1998, he explained that the network was doing two things. Like books, it could accumulate knowledge. But, more importantly, it presented knowledge in a way that connected it with other information whereas, in a book, information stayed isolated.

Because the web was easy to use with hyperlinks going from one document to the next, the internet could now be used by anyone, and not only by computer literate users. There were 100 million internet users in

December 1997, with one million new users per month, and 300 million users in December 2000.

A worldwide expansion

North America was leading the way in computer science and communication technology, with significant funding and cheap computers compared to Europe. A connection to the internet was much cheaper too.

In some European countries, internet users needed to surf the web at night (including the author of these lines), when phone rates by the minute were cheaper, to cut their expenses. In late 1998 and early 1999, some users in France, Germany and Italy launched a movement to boycott the internet one day per week, as a way to force internet providers and phone companies to set up a special monthly rate. This action paid off, and providers began to offer "internet rates".

In summer 1999, the number of internet users living outside the U.S. reached 50%.

In summer 2000, the number of internet users having a mother tongue other than English also reached 50%, and went on steadily increasing then. According to statistics regularly published on the website of Global Reach, a marketing consultancy promoting internationalization and localization, they were 52.5% in summer 2001, 57% in December 2001, 59.8% in April 2002, 64.4% in September 2003 (including 34.9% non-English-speaking Europeans and 29.4% Asians), and 64.2% in March 2004 (including 37.9% non-English-speaking Europeans and 33% Asians).

Broadband became the norm over the years. Jean-Paul, webmaster of the hypermedia website cotres.net, summarized things in January 2007: "I feel that we are experiencing a 'floating' period between the heroic ages, when

we were moving forward while waiting for the technology to catch up, and the future, when high-speed broadband will unleash forces that just begin to move, for now only in games.”

The internet of the future

The internet of the future could be a “pervasive” network allowing us to connect in any place and at any time on any device through a single omnipresent network.

The concept of a “pervasive” network was developed by Rafi Haladjian, founder of the European company Ozone, who explained on its website in 2007 that “the new wave would affect the physical world, our real environment, our daily life in every moment. We will not access the network any more, we will live in it. The future components of this network (wired parts, non wired parts, operators) will be transparent to the final user. The network will always be open, providing a permanent connection anywhere. It will also be agnostic in terms of applications, as a network based on the internet protocols themselves.” We do look forward to this.

As for the content of the internet, Timothy Leary, a visionary writer, described it in 1994 in his book “Chaos & Cyber Culture” as gigantic glass towers containing the whole world information, with free access, through the cyberspace, not only to all books, but also to all pictures, all movies, all TV shows, and all other data. In 2011, we are not there yet, but we are getting there.

1990 > THE INVENTION OF THE WEB

[Summary] The World Wide Web was invented in 1990 by Tim Berners-

Lee at CERN (European Center for Nuclear Research), Geneva, Switzerland. In 1989, Tim Berners-Lee networked documents using hypertext. In 1990, he developed the first HTTP (HyperText Transfer Protocol) server and the first web browser. In 1991, the web was operational and radically changed the way people were using the internet. Hypertext links allowed us to move from one textual or visual document to another with a simple click of the mouse. Information became interactive, thus more attractive to many users. Later on, this interactivity was further enhanced with hypermedia links that could link texts and images with video and sound. The World Wide Web Consortium (W3C) was founded in October 1994 to develop protocols for the web.

The World Wide Web was invented in 1990 by Tim Berners-Lee, a researcher at CERN (European Center for Nuclear Research), Geneva, Switzerland, who made the internet accessible to all.

How the web started

In 1989, Tim Berners-Lee networked documents using hypertext. In 1990, he developed the first HTTP (HyperText Transfer Protocol) server and the first web browser. In 1991, the web was operational and made the internet accessible to all. Hypertext links allowed us to move from one textual or visual document to another with a simple click of the mouse. Information became interactive, thus more attractive to many users. Later on, this interactivity was further enhanced with hypermedia links that could link texts and images with video and sound.

Developed by NCSA (National Center for Supercomputing Applications) at the University of Illinois (USA) and distributed free of charge in November 1993, Mosaic was the first browser for the general public, and contributed greatly to the development of the web. In early 1994, part of the

Mosaic team migrated to the Netscape Communications Corporation to develop a new browser called Netscape Navigator. In 1995, Microsoft launched its own browser, the Internet Explorer. Other browsers were launched then, like Opera and Safari, Apple's browser.

The World Wide Web Consortium (W3C) was founded in October 1994 to develop interoperable technologies (specifications, guidelines, software, other tools) for the web, for example specifications for markup languages (HTML, XML and others). It also acted as a forum for information, commerce, communication and collective understanding. In 1998, the section Internationalization/Localization gave access to some protocols for creating a multilingual website: HTML, base character set, new tags and attributes, HTTP, language negotiation, URLs and other identifiers including non-ASCII characters, etc.

Tim Berners-Lee's dream

Pierre Ruetschi, a journalist for the Swiss daily "Tribune de Genève", asked Tim Berners-Lee on 20 December 1997: "Seven years later, are you satisfied with the way the web has evolved?". He answered that, if he was pleased with the richness and diversity of information, the web still lacked the power planned in its original design. He would like "the web to be more interactive, and people to be able to create information together", and not only to be information consumers. The web was supposed to become a "medium for collaboration, a world of knowledge that we share."

In an essay posted on his webpage, Tim Berners-Lee wrote in May 1998: "The dream behind the web is of a common information space in which we communicate by sharing information. Its universality is essential: the fact that a hypertext link can point to anything, be it personal, local or global, be it draft or highly polished. There was a second part of the dream, too, dependent on the web being so generally used that it became a realistic mirror

(or in fact the primary embodiment) of the ways in which we work and play and socialize. That was that once the state of our interactions was online, we could then use computers to help us analyze it, make sense of what we are doing, where we individually fit in, and how we can better work together." (excerpt from "The World Wide Web: A very short personal history")

The web 2.0

According to Netcraft, a company tracking data on the internet, the number of websites went from one million (April 1997) to 10 million (February 2000), 20 million (September 2000), 30 million (July 2001), 40 million (April 2003), 50 million (May 2004), 60 million (March 2005), 70 million (August 2005), 80 million (April 2006), 90 million (August 2006) and 100 million (November 2006), with a growing number of personal websites and blogs.

The term “web 2.0” was invented in 2004 by Tim O’Reilly, a publisher of computer books, as a title for a series of conferences he was organizing. The web 2.0 may begin to answer Tim Berners-Lee’s dream as a web based on community and sharing, with many collaborative projects across borders and languages.

Fifteen years after the invention the web, Wired stated in its August 2005 issue that less than half of the web was commercial, with the other half being run by passion. As for the internet, according to the French daily Le Monde dated 19 August 2005, its three powers — ubiquity, variety and interactivity — made its potential use quasi infinite.

Robert Beard, a language teacher at Bucknell University, Pennsylvania, and the founder of A Web of Online Dictionaries in 1995, wrote as early as September 1998: "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."

1990 > THE LINGUIST LIST

[Summary] The LINGUIST List was founded by Anthony Rodrigues Aristar in 1990 at the University of Western Australia, with 60 subscribers, before moving to Texas A&M University in 1991, with Eastern Michigan University established as the main editing site for the list. In 1997, emails sent to the distribution list were also available on the list's own website, in several sections: the profession (conferences, linguistic associations, programs), 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). The LINGUIST List is a component of the WWW Virtual Library for linguistics.

The LINGUIST List was founded by Anthony Rodrigues Aristar in 1990 at the University of Western Australia, as a mailing list for academic linguists.

With 60 subscribers, it moved to Texas A&M University in 1991, with Eastern Michigan University being established as the main editing site for the list.

In 1997, emails sent to the distribution list were also available on the list's own website, in several sections: the profession (conferences, linguistic

associations, programs), 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). The LINGUIST List is a component of the WWW Virtual Library for linguistics.

Helen Dry, co-moderator of the LINGUIST List since 1991, wrote in August 1998: "The LINGUIST List, which I moderate, has a policy of posting in any language, since it is 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 five 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."

She added in July 1999: "We are beginning to collect some primary data. For example, we have searchable databases of dissertation abstracts relevant to linguistics, of information on graduate and undergraduate linguistics programs, and of professional information about individual linguists. The dissertation abstracts collection is, to my knowledge, the only freely available electronic compilation in existence.

1991 > FROM ASCII TO UNICODE

[Summary] Used since the beginning of computing, ASCII (American

Standard Code for Information Interchange) is a 7-bit coded character set for information interchange in English. It was published in 1963 by ANSI (American National Standards Institute). With the internet spreading worldwide, to communicate in English (and Latin) was not enough anymore. The accented characters of several European languages and characters of some other languages were taken into account from 1986 onwards with 8-bit variants of ASCII, also called extended ASCII, that provided sets of 256 characters. But problems were not over until the publication of Unicode in January 1991 as a new universal encoding system. Unicode provided "a unique number for every character, no matter what the platform, no matter what the program, no matter what the language", and could handle 65,000 characters or ideograms.

With the internet spreading worldwide, the use of ASCII and extended ASCII was not enough anymore, thus the need to take into account all languages with Unicode, whose first version was published in January 1991.

Used since the beginning of computing, ASCII (American Standard Code for Information Interchange) is a 7-bit coded character set for information interchange in English (and Latin). It was published in 1963 by ANSI (American National Standards Institute). The 7-bit plain ASCII, also called Plain Vanilla ASCII, is a set of 128 characters with 95 printable unaccented characters (A-Z, a-z, numbers, punctuation and basic symbols), the ones that are available on the American / English keyboard.

With computer technology spreading outside North America, the accented characters of several European languages and characters of some other languages were taken into account from 1986 onwards with 8-bit variants of ASCII, also called extended ASCII, that provided sets of 256 characters.

Brian King, director of the WorldWide Language Institute (WWLI), explained in September 1998: “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 non-alphabetic systems like Chinese. But tradition has been turned upside down. Technology has been popularized. (...)”

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. (...)”

The most recent development [in 1998] is Unicode. Although still evolving and only just being incorporated into the latest software, this new coding system translates each character into 16 bits. Whereas 8-bit 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 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."

First published in January 1991, Unicode "provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language" (excerpt from the website). This double-byte platform-independent encoding provides a basis for the processing, storage and interchange of text data in any language. Unicode is maintained by the Unicode Consortium, with its variants UTF-8, UTF-16 and UTF-32 (UTF: Unicode Transformation Format), and is a component of the specifications of the World Wide Web Consortium (W3C). Unicode has replaced ASCII for text files on Windows platforms since 1998. Unicode surpassed ASCII on the internet in December 2007.

1994 > TRAVLANG, TRAVEL AND LANGUAGES

[Summary] Travlang was the first website to offer links to free basic translation dictionaries, intended for travelers and the general public. As a first step, Michael C. Martin created in 1994 a "Foreign Languages for Travelers" section on his university website when he was a physics student in New York. One year later, he launched Travlang, a site that quickly became a major portal for travel and languages, and won a best travel site award in 1997. Travlang was still maintained in 1998 by Michel C. Martin, now an researcher in experimental physics at the Lawrence Berkeley National

Laboratory in California. The section “Translating Dictionaries” gave access to free basic online dictionaries in a number of languages (Afrikaans, Czech, Danish, Dutch, Esperanto, Finnish, French, Frisian, German, Hungarian, Italian, Latin, Norwegian, Portuguese, Spanish, Swedish). Other sections offered links to language dictionaries, translation services and language schools.

Travlang was the first website to offer links to free basic translation dictionaries, intended for travelers and the general public.

As a first step, Michael C. Martin created in 1994 a “Foreign Languages for Travelers” section on his university website when he was a physics student in New York. One year later, he launched Travlang, a site that quickly became a major portal for travel and languages, and won a best travel site award in 1997.

Travlang was still maintained in 1998 by Michel C. Martin, now an researcher in experimental physics at the Lawrence Berkeley National Laboratory in California.

The section “Foreign Languages for Travelers” gave links to online tools to learn 60 languages. The section “Translating Dictionaries” gave access to free basic online dictionaries in a number of languages (Afrikaans, Czech, Danish, Dutch, Esperanto, Finnish, French, Frisian, German, Hungarian, Italian, Latin, Norwegian, Portuguese, Spanish, Swedish). Other sections offered links to translation services, language schools and multilingual bookstores. People could also book their hotel, car or plane ticket, look up exchange rates and browse an index of 7,000 links to other language and travel sites.

Michael C. Martin wrote in August 1998: "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. (...) 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."

What about the future? "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."

Michael C. Martin sold Travlang to GourmetMarket.com in February 1999. GourmetMarket.com sold it to iiGroup in January 2000. By July 2000, the site was pulling in two million visitors a month.

1995 > THE INTERNET DICTIONARY PROJECT

[Summary] Tyler Chambers first created the Human-Languages Page (H-LP) in May 1994 as an index of language-related internet resources in a number of languages. In 1995, Tyler launched a second project, the Internet Dictionary Project (IDP), as a collaborative project to create free online dictionaries from English to other languages (French, German, Italian, Latin, Portuguese, Spanish). As explained in 1998 on the project's 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."

In 1995, Tyler Chambers launched the Internet Dictionary Project (IDP) as a collaborative project to create free online dictionaries from English to other languages (French, German, Italian, Latin, Portuguese, Spanish).

Before launching the Internet Dictionary Project, Tyler created the Human-Languages Page (H-LP) in May 1994 as an index of linguistic internet resources. In 1998, there were 1,800 language-related resources in 100 languages, with six subject listings (languages and literature, schools and institutions, linguistics resources, products and services, organizations, jobs and internships) and two category listings (dictionaries, language lessons).

What exactly was the Internet Dictionary Project? As explained in 1998 on the project's 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 online 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 online 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 wrote in September 1998 in an email interview: "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 online. This was back before Netscape even existed — Mosaic was almost the only web browser, and webpages 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 online. (...)

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. (...) 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? (...)

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. (...) Language will become even more important than it

already is when the entire planet can communicate with everyone else (via the web, chat, games, email, and whatever future applications haven't even been invented yet)."

In spring 2001, the Human-Languages Page merged with the Languages Catalog, a section of the WWW Virtual Library, to become iLoveLanguages. In September 2003, iLoveLanguages provided an index of 2,000 linguistic resources in 100 languages. As for the Internet Dictionary Project, Tyler ran out of time to maintain it and removed the ability to update the dictionaries in January 2007. People can still search the available dictionaries or download the archived files.

1995 > NETGLOS, A GLOSSARY OF THE INTERNET

[Summary] Launched in 1995 by the WorldWide Language Institute (WWLI), an institute providing language instruction via the internet, NetGlos — which stands for "Multilingual Glossary of Internet Terminology" — was compiled as a voluntary collaborative project by a number of translators and other language professionals worldwide. In September 1998, NetGlos was available in 13 languages (Chinese, Croatian, English, Dutch/Flemish, French, German, Greek, Hebrew, Italian, Maori, Norwegian, Portuguese, Spanish). As explained by Brian King, director of the WorldWide Language Institute, in September 1998: “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 loan word 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 non technical user.”

NetGlos — which stands for "Multilingual Glossary of Internet Terminology" — was compiled as a voluntary collaborative project by a number of translators and other language professionals worldwide.

NetGlos was launched in 1995 by the WorldWide Language Institute (WWLI), an institute providing language instruction via the internet. Three years later, NetGlos was available in 13 languages (Chinese, Croatian, English, Dutch/Flemish, French, German, Greek, Hebrew, Italian, Maori, Norwegian, Portuguese, Spanish).

As explained by Brian King, director of the WorldWide Language Institute, in September 1998: "Much of the technical terminology on the web is still not translated into other languages. And as we found with (...) 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 loan word 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 non technical user. The latest version of NetGlos is the Russian one and it should be available in a couple of weeks or so [in late September 1998]. It will no doubt be an excellent example of the ongoing, dynamic process of 'russification' of web terminology."

How about the future? "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. Lifelong 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."

1995 > VARIOUS LANGUAGES ON OUR SCREEN

[Summary] In December 1995, Yoshi Mikami, a computer scientist at Asia Info Network in Fujisawa, Japan, created the website "The Languages of the World by Computers and the Internet", also known as the Logos Home Page or Kotoba Home Page, "to summarize there the brief history, linguistic and phonetic features, writing system and computer processing aspects for each of the six major languages of the world, in English and Japanese". As a second step, Yoshi Mikami was also the co-author (with Kenji Sekine and Nobutoshi Kohara) of "The Multilingual Web Guide" (Japanese edition), a print book published in August 1997 by O'Reilly Japan, and translated in 1998 into English, French and German.

In December 1995, Yoshi Mikami created the website "The Languages of the World by Computers and the Internet", also known as the Logos Home

Page or Kotoba Home Page, "to summarize there the brief history, linguistic and phonetic features, writing system and computer processing aspects for each of the six major languages of the world, in English and Japanese".

Yoshi Mikami was a computer scientist at Asia Info Network in Fujisawa, Japan. As a second step, one year after launching his website, he was also the co-author (with Kenji Sekine and Nobutoshi Kohara) of "The Multilingual Web Guide" (Japanese edition), a print book published in August 1997 by O'Reilly Japan, and translated in 1998 into English, French and German.

Yoshi explained in December 1998: "My native tongue is Japanese. Because I had my graduate education in the U.S. and worked in the computer business, I became bilingual in Japanese and American English. I was always interested in languages and different 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 aspects 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 help me write a book on viewing, understanding and creating multilingual webpages, which was published in August 1997 as 'The Multilingual Web Guide', in a Japanese edition, the world's first book on such a subject."

As for multilingualism, Yoshi added: "Thousands of years ago, in Egypt, China and elsewhere, people were more concerned about communicating their laws and thoughts not in just one language, but in several. In our modern world, most nation states have each adopted one language for their own use. I predict greater use of different languages and multilingual pages on the internet, not a simple gravitation to American English, and also more creative use of multilingual computer translation. 99% of the websites created

in Japan are written in Japanese.”

1995 > GLOBAL REACH, PROMOTING LOCALIZATION

[Summary] Ten years after founding Euro-Marketing Associates, a company based in San Francisco and Paris, Bill Dunlap launched in 1995 Global Reach, a marketing consultancy helping U.S. companies to expand their internet presence into an international framework. This included translating a website into other languages, actively promoting it, and using local online banner advertising to increase local website traffic. Bill Dunlap explained in December 1998: “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.”

In 1995, Bill Dunlap launched Global Reach, a marketing consultancy helping U.S. companies to expand their internet presence into an international framework. This included translating a website into other languages, actively promoting it, and using local online banner advertising to increase local website traffic.

Ten years earlier, Bill Dunlap founded Euro-Marketing Associates, a company based in San Francisco and Paris. He wrote in December 1998: “There are so few people in the U.S. interested in communicating in many languages — most Americans are still under the delusion that the rest of the world speaks English. However, in Europe, the countries are small enough so

that an international perspective has been necessary for centuries. (...)

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 have turned these activities to their online dimension, and have come to champion European e-commerce among my fellow American compatriots. 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. (...)

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 online international market seriously. Global Reach is a means for you to extend your website to many countries, speak to online visitors in their own language and reach online markets there."

Bill added in July 1999: "After a website's home page is available in several languages, the next step is the development of content in each language. A webmaster will notice which languages draw more visitors (and sales) than others, and these are the places to start in a multilingual web promotion campaign. At the same time, it is always good to increase the number of languages available on a website: just a home page translated into other languages would do for a start, before it becomes obvious that more should be done to develop a certain language branch on a website."

1996 > ONELOOK DICTIONARIES, A “FAST FINDER”

[Summary] Robert Ware launched OneLook Dictionaries in April 1996 as a "fast finder" in hundreds of online dictionaries covering various topics: business, computer/internet, medical, miscellaneous, religion, science, sports, technology, general and slang. He wrote in September 1998: "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." OneLook Dictionaries could browse 2 million words from 425 dictionaries in 1998, 2.5 million words from 530 dictionaries in 2000, 5 million words from 910 dictionaries in 2003, and 19 million words from 1,060 dictionaries in 2010.

Robert Ware launched OneLook Dictionaries in April 1996 as a "fast finder" in hundreds of online dictionaries covering various topics: business, computer/internet, medical, miscellaneous, religion, science, sports, technology, general and slang.

He wrote in September 1998: "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."

Robert Ware also wrote about a personal experience showing the internet

could promote both a common language and multilingualism: "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 email. 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."

OneLook Dictionaries could browse 2 million words from 425 dictionaries in 1998, 2.5 million words from 530 dictionaries in 2000, 5 million words from 910 dictionaries in 2003, and 19 million words from 1,060 dictionaries in 2010.

1997 > 82.3% OF THE WEB IN ENGLISH

[Summary] The internet was born in 1974 in the U.S. before spreading to the English-speaking community and then worldwide. This explain why it took a little while for other languages than English to be distributed. The first major study about language distribution on the web was run by Babel, a joint project from the Internet Society and Alis Technologies to contribute to the internationalization of the internet. The results were published in June 1997 in seven languages on a webpage named "Web Languages Hit Parade". The main languages available on the web were English with 82.3%, German with 4.0%, Japanese with 1.6%, French with 1.5%, Spanish with 1.1%, Swedish with 1.1%, and Italian with 1.0%. Three years later, in spring 2000, non-English-speaking internet users reached 50%, with a percentage steadily increasing then.

The first major study about language distribution on the web was run in 1997 by Babel, a joint project from the Internet Society and Alis Technologies to contribute to the internationalization of the internet.

The internet was born in 1974 in the U.S. before spreading to the English-speaking community and then worldwide. This explain why it took a little while for other languages than English to be distributed. People from all over the world began to have access to the internet, despite a connection that was far from cheap in a number of countries, and to post webpages in their own languages. The percentage of webpages in English slowly decreased from nearly 100% in 1983 to 85% in 1997.

“Towards communicating on the internet in any language...” was the subtitle of Babel, launched in 1997 as a plurilingual website in seven languages (English, French, German, Italian, Portuguese, Spanish, Swedish), with information about the world's languages and a typographical and linguistic glossary. A section named "The Internet and Multilingualism" gave

information on how to develop a multilingual website, and how to code the "world's writing".

Babel ran the first major study relating to distribution of languages on the web. The results were published in June 1997 in seven languages on a webpage named "Web Languages Hit Parade". The main languages of the web were English with 82.3%, German with 4.0%, Japanese with 1.6%, French with 1.5%, Spanish with 1.1%, Swedish with 1.1%, and Italian with 1.0%.

According to Randy Hobler, a consultant in internet marketing for language translation software and services, interviewed in September 1998: "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."

Randy also explained in the same email interview: "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 U.S., as well as odd places like Spanish-speaking Morocco."

According to Global Reach, a marketing consultancy promoting localization, there were 56 million non-English-speaking users in July 1998, with 22.4% Spanish-speaking users, 12.3% Japanese-speaking users, 14% German-speaking users and 10% French-speaking users. 15% of Europe's half a billion population spoke English as a first language, 28% didn't speak

English at all, and 32% were using the web in English.

In summer 1999, the number of internet users living outside the U.S. reached 50%.

In summer 2000, the number of internet users having a mother tongue other than English also reached 50%, and went on steadily increasing then. According to statistics regularly published online by Global Reach, they were 52.5% in summer 2001, 57% in December 2001, 59.8% in April 2002, 64.4% in September 2003 (including 34.9% non-English-speaking Europeans and 29.4% Asians), and 64.2% in March 2004 (including 37.9% non-English-speaking Europeans and 33% Asians).

1997 > THE INTERNET, A TOOL FOR MINORITY LANGUAGES

[Summary] Despite the so-called hegemony of the English language, the internet was also a good tool for minority languages, as stated by Caoimhín Ó Donnaíle, who has taught computing at the Institute Sabhal Mòr Ostaig, on the Isle of Skye, in Scotland. Caoimhín has maintained the trilingual (Scottish Gaelic, Irish Gaelic, English) college website, as the main site worldwide with information on Scottish Gaelic, with a trilingual list of European minority languages. The internet could be a tool to develop a "cultural identity" for any language, while using the English language for this, as stated by Guy Antoine, who founded Windows on Haiti in April 1998 to promote the Haitian culture and language.

Despite the so-called hegemony of the English language, the internet was also a good tool for minority languages, as stated by

Caoimhín Ó Donnaíle, who has taught computing at the Institute Sabhal Mòr Ostaig, on the Isle of Skye, in Scotland.

Caoimhín has maintained the trilingual (Scottish Gaelic, Irish Gaelic, English) college website, as the main site worldwide with information on Scottish Gaelic, with a trilingual list of European minority languages.

Interviewed in August 1998, Caoimhín saw four main points for the growth of a multilingual web: “(a) The internet has contributed and will contribute to the wildfire spread of English as a world language. (b) 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. (c) The web is very useful for delivering language lessons, and there is a big demand for this. (d) The Unicode (ISO 10646) character set standard is very important and will greatly assist in making the Internet more multilingual.”

How about the Gaelic language? Caoimhín wrote in May 2001: "Students do everything by computer, use Gaelic spell-checking, a Gaelic online terminology database. There are more hits on our website. There is more use of sound. Gaelic radio (both Scottish and Irish) is now available continuously worldwide via the internet. A major project has been the translation of the Opera web-browser into Gaelic — the first software of this size available in Gaelic."

What about endangered languages? "I would emphasize the point that as regards the future of endangered languages, the internet speeds everything up. If people don't care about preserving languages, the internet and accompanying globalization will greatly speed their demise. If people do care about preserving them, the internet will be a tremendous help."

Robert Beard, co-founder of the web portal yourDictionary.com, wrote in

January 2000: "While English still dominates the web, the growth of monolingual non-English websites is gaining strength with the various solutions to the font problems. Languages that are endangered are primarily languages without writing systems at all (only 1/3 of the world's 6,000+ languages have writing systems). I still do not see the web contributing to the loss of language identity and still suspect it may, in the long run, contribute to strengthening it. More and more Native Americans, for example, are contacting linguists, asking them to write grammars of their language and help them put up dictionaries. For these people, the web is an affordable boon for cultural expression."

The internet could be a tool to develop a "cultural identity" for any language, while using the English language for this, as stated by Guy Antoine, who founded Windows on Haiti in April 1998 to promote the Haitian culture and language.

Guy wrote in November 1999: "In Windows on Haiti, the primary language of the site is English, but one will equally find a center of lively discussion conducted in 'Kreyòl'. In addition, one will find documents related to Haiti in French, in the old colonial Creole, and I am open to publishing others in Spanish and other languages. I do not offer any sort of translation, but multilingualism is alive and well at the site, and I predict that this will increasingly become the norm throughout the web. (...)

The internet can serve, first of all, as a repository of useful information on minority languages that might otherwise vanish without leaving a trace. Beyond that, I believe that it provides an incentive for people to learn languages associated with the cultures about which they are attempting to gather information. One soon realizes that the language of a people is an essential and inextricable part of its culture. (...) 'Kreyòl' (Creole for the non-initiated) is primarily a spoken language, not a widely written one. I see

the web changing this situation more so than any traditional means of language dissemination."

Guy added in June 2001: "Kreyòl is the only national language of Haiti, and one of its two official languages, the other being French. It is hardly a minority language in the Caribbean context, since it is spoken by eight to ten million people. (...) I have taken the promotion of Kreyòl as a personal cause, since that language is the strongest of bonds uniting all Haitians. (...) I have created two discussion forums on my website Windows on Haiti, held exclusively in Kreyòl. One is for general discussions on just about everything but obviously more focused on Haiti's current socio-political problems. The other is reserved only to debates of writing standards for Kreyòl. Those debates have been quite spirited and have met with the participation of a number of linguistic experts. The uniqueness of these forums is their non-academic nature."

1997 > A EUROPEAN TERMINOLOGY DATABASE

[Summary] Launched in 1997 by the Translation Service of the European Commission, Eurodicautom was a multilingual terminology database of economic, scientific, technical and legal terms and expressions, with language pairs for the eleven official languages of the European Union (Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, Swedish), and Latin. There were 120,000 daily visits on average in 2003. In late 2003, Eurodicautom announced its integration into a larger terminology database in partnership with other institutions of the European Union. The new database, called IATE (InterActive Terminology for Europe), would be available in more than 20 languages, because of the enlargement of the European Union planned in 2004. IATE was launched on

the intranet of some European institutions in spring 2004 and on the internet for free in March 2007.

Eurodicautom was a multilingual terminology database of economic, scientific, technical and legal terms and expressions, with language pairs for the eleven official languages of the European Union, and Latin.

Eurodicautom was initially developed to assist in-house translators. A free online version was available on the web in 1997 for European Union officials and for language professionals throughout the world.

Eurodicautom covered "a broad spectrum of human knowledge", mainly relating to economy, science, technology and legislation in the European Union (EU), to answer the needs of the 15 member countries in 11 official languages (Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, Swedish), plus Latin.

The project of a larger terminology database was studied as early as 1999 to merge the existing databases for a better inter- institutional cooperation between the European organizations. The project partners were the European Commission, the European Parliament, the Council of the European Union, the Court of Justice, the European Court of Auditors, the European Economic and Social Committee, the Committee of the Regions, the European Investment Bank, the European Central Bank, and the Translation Centre for the Bodies of the European Union.

Eurodicautom had 12,000 visits a day in late 2003, when it closed to prepare for a larger terminology database that would include the databases of other official European institutions. The new database would be available in many more languages, more than 20 languages instead of 12, because of the

Enlargement of the European Union planned in 2004 to include new countries from Central and Eastern Europe. The European Union went from 15 country members to 25 country members in May 2004, and 27 country members in January 2007.

IATE (InterActive Terminology for Europe) was launched in March 2007 as an eagerly free service on the web, after been launched in summer 2004 on the intranet of the participating European institutions, with 1.4 million entries in the 23 official languages of the European Union (Bulgarian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovene, Spanish, Swedish), plus Latin.

The website has been maintained by the Translation Center of the European Union institutions in Luxembourg. According to the IATE brochure, also available in the 23 official languages, IATE offered 8,4 million words in 2010, including 540,000 abbreviations and 130.000 expressions.

1997 > BABEL FISH, A FREE TRANSLATION SOFTWARE

[Summary] In December 1997, the search engine AltaVista launched the first free machine translation software called Babel Fish or AltaVista Translation, which could translate webpages or short texts from English into French, German, Italian, Portuguese or Spanish, and vice versa. The software was developed by Systran (an acronym for "System Translation"), a company specializing in automated language solutions. Babel Fish was a “hit” among the 12 million internet users of the time, who included more and more non-English-speaking users, and greatly contributed to a plurilingual web. Other tools were developed then by Alis Technologies, Globalink, Lernout &

Hauspie and Softissimo, with free and/or paid versions available on the web.

In December 1997, the search engine AltaVista launched Babel Fish as the first free machine translation software from English to five other languages.

At the time, the interface of Yahoo! was available in seven languages (English, French, German, Japanese, Korean, Norwegian, Swedish), to take into account a growing number of non-English-speaking users. When a search didn't give any result in Yahoo!, it was automatically shunted to AltaVista, and vice versa.

Babel Fish, also called AltaVista Translation, could translate webpages from English into French, German, Italian, Portuguese or Spanish, and vice versa, the original page and the translation being face-to-face on the screen. Translating any short text was also possible with a "copy and paste". The result was far from perfect but helpful, as well as instantaneous and free unlike a high-quality professional translation. Non-English-speaking users were thrilled. Babel Fish greatly contributed to a plurilingual web.

Backed up by plurilingual dictionaries with 12.5 million entries, Babel Fish was developed by Systran (an acronym for "System Translation"), a company specializing in automated language solutions. As explained on Systran's 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."

Machine translation was defined as such on the website of the European

Association for Machine Translation (EAMT): "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."

Other translation software was developed then by Alis Technologies, Globalink, Lernout & Hauspie and Softissimo, with paid and/or free versions available on the web. As for Babel Fish, it moved to Yahoo!'s website in May 2008.

1997 > THE TOOLS OF THE TRANSLATION COMPANY LOGOS

[Summary] In December 1997, Logos, a global translation company based in Modena, Italy, decided to put on the web for free the professional tools used by its translators, for the internet community to be able to use them as well. These tools were the Logos Dictionary, a multilingual dictionary with 7.5 billion words (in fall 1998); the Logos Wordtheque, a multilingual library with 328 billion words extracted from translated novels, technical manuals, and other texts; the Logos Linguistic Resources, a database of 553 glossaries; and the Logos Universal Conjugator, a database for verbs in 17 languages. In 2007, the Logos Library (formerly Wordtheque) included 710 billion words, Linguistic Resources (no change of name) included 1,215 glossaries, and the Universal Conjugator (formerly Conjugation of Verbs) included verbs in 36

languages.

In December 1997, Logos, a global translation company, decided to put on the web all the professional tools used by its translators, for the internet community to freely use them as well.

Logos was founded by Rodrigo Vergara in 1979, with headquarters in Modena, Italy. In 1997, Logos had 300 in-house translators and 2,500 free-lance translators worldwide, who processed around 200 texts per day.

The linguistic tools available online were the Logos Dictionary, a multilingual dictionary with 7.5 billion words (in fall 1998); the Logos Wordtheque, a multilingual library with 328 billion words extracted from translated novels, technical manuals, and other texts, that could be searched by language, word, author or title; the Logos Linguistic Resources, a database of 500 glossaries; and the Logos Universal Conjugator, a database for verbs in 17 languages.

When interviewed by Annie Kahn in an article of the French daily *Le Monde* dated 7 December 1997, Rodrigo Vergara, 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. This move has in fact attracted many customers, and also allowed us to widen our network of translators, thanks to contacts made in the wake of the initiative."

In the same article, called "Les mots pour le dire" (The words to tell it), Annie Kahn wrote: "The Logos site is much more than a mere dictionary or a collection of links to other online dictionaries. The cornerstone is the

document search program, 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, including in foreign translations, thanks to a partnership agreement with the famous online bookstore Amazon.com. However, if no text containing the required word is found, the program acts as a search engine, sending the user to other web sources containing this word. 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 suggestions, after which Logos translators check the suggested translations they receive."

Ten years later, in 2007, the Logos Library (formerly Wordtheque) included 710 billion words, Linguistic Resources (no change of name) included 1,215 glossaries, and the Universal Conjugator (formerly Conjugation of Verbs) included verbs in 36 languages.

1997 > SPECIALIZED TERMINOLOGY DATABASES

[Summary] Some international organizations have run terminology databases in their own field of expertise for their translation services. In 1997, some databases were freely available on the web, to be used by language professionals throughout the world and by the internet community at large, for example ILOTERM, maintained by the International Labor Organization (ILO), TERMITE (ITU Telecommunication Terminology Database), maintained by the International Telecommunication Union (ITU), and WHOTERM (WHO Terminology Information System), maintained by

the World Health Organization (WHO).

In 1997, some specialized terminology databases maintained by international organizations in their own field of expertise were freely available on the web, to be used by language professionals throughout the world and by the internet community at large.

Here are three examples with ILOTERM, maintained by the International Labor Organization (ILO), TERMITE (ITU Telecommunication Terminology Database), maintained by the International Telecommunication Union (ITU), and WHOTERM (WHO Terminology Information System), maintained by the World Health Organization (WHO).

ILOTERM is a quadrilingual (English, French, German, Spanish) terminology database maintained by the Terminology and Reference Unit of the Official Documentation Branch (OFFDOC) at the International Labor Office (ILO) in Geneva, Switzerland. As explained on its website, the primary purpose of ILOTERM is to provide solutions, reflecting current usage, to terminology issues in the social and labor fields. Terms are available in English with their French, Spanish and German equivalents. The database also includes the ILO structure and programs, official names for international institutions, national bodies and employers' and workers' organizations, and names of international meetings and symposiums.

TERMITE, which stands for “Telecommunication Terminology Database”, is a quadrilingual (English, French, Spanish, Russian) terminology database maintained by the Terminology, References and Computer Aids to Translation Section of the Conference Department at the International Telecommunication Union (ITU) in Geneva, Switzerland. This database has

been built on the content of all ITU printed glossaries since 1980, and regularly updated with recent entries.

WHOTERM, which stands for “WHO Terminology Information System”, is a trilingual (English, French, Spanish) database maintained by the World Health Organization (WHO) in Geneva, Switzerland. It has included: (a) the WHO General Dictionary Index in English, with the French and Spanish equivalents; (b) three glossaries in English: Health for All, Programme Development and Management, and Health Promotion; (c) the WHO TermWatch, an awareness service from the Technical Terminology, reflecting the current WHO usage, but not necessarily terms officially approved by WHO, with links to health-related terminology.

1998 > THE NEED FOR A “LINGUISTIC DEMOCRACY”

[Summary] Brian King, director of the WorldWide Language Institute (WWLI), brought up the concept of "linguistic democracy" in September 1998: "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."

Brian King, director of the WorldWide Language Institute (WWLI), brought up the concept of "linguistic democracy" in September 1998:

"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."

For Brian King, one factor contributing to the development of a multilingual internet is the "competition for a chunk of the 'global market' by major industry players", with "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."

Another factor is the development of 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!"

Founder of Euro-Marketing Associates and its virtual branch Global Reach, Bill Dunlap championed the assets of e-commerce in Europe among his fellow compatriots in the U.S., promoting the internationalization and localization of their websites. He wrote in December 1998: "There are so few people in the U.S. interested in communicating in many languages — most Americans are still under the delusion that the rest of the world speaks English. However, in Europe, the countries are small enough so that an international perspective has been necessary for centuries."

Peter Raggett, deputy-head (and then head) of the Central Library of OECD (Organization for Economic Cooperation and Development), wrote in August 1999: "I think it is incumbent on European organizations and businesses to try and offer websites in three or four languages if resources permit. In this age of globalization and electronic commerce, businesses are finding that they are doing business across many countries. Allowing French, German, Japanese speakers to easily read one's website as well as English speakers will give a business a competitive edge in the domain of electronic trading."

As the internet quickly spread worldwide, companies needed to offer bilingual, trilingual, even plurilingual websites to reach as large an audience as possible, while adapting their content to a given audience, either a country or a linguistic community. Thus the need to internationalize and localize websites, which became a major trend in the late 1990s and early 2000s, with English- language companies and organizations setting up plurilingual websites, in English and other languages, and non-English-language companies and organizations setting up websites in their own language(s) and English.

1999 > BILINGUAL DICTIONARIES IN WORDREFERENCE.COM

[Summary] Michael Kellogg created WordReference.com in 1999. He wrote much later on his website: "I started this site in 1999 in an effort to provide free online bilingual dictionaries and tools to the world. The site has grown gradually ever since to become one of the most- used online dictionaries, and the top online dictionary for its language pairs of English-Spanish, English-French, English-Italian, Spanish-French, and Spanish-Portuguese. It is consistently ranked in the top 500 most-visited websites in the world. I am proud of my history of innovation with dictionaries on the internet. Many of the features such as being able to click any word in a dictionary entry were first implemented by me." WordReference was also provided high-quality language forums, and lighter versions of some dictionaries for mobile devices.

Michael Kellogg created WordReference.com in 1999 to offer free online bilingual translation dictionaries.

Much later, Michael wrote on his website: "I started this site in 1999 in an effort to provide free online bilingual dictionaries and tools to the world. The site has grown gradually ever since to become one of the most-used online dictionaries, and the top online dictionary for its language pairs of English-Spanish, English-French, English-Italian, Spanish-French, and Spanish-Portuguese. It is consistently ranked in the top 500 most-visited websites in the world. I am proud of my history of innovation with dictionaries on the internet. Many of the features such as being able to click any word in a dictionary entry were first implemented by me."

How was the idea behind his project? "The internet has done an incredible

job of bringing the world together in the last few years. Of course, one of the greatest barriers has been language. Much of the content is in English and many, many users are reading English-language webpages as a second language. I know from my own experiences with Spanish-language websites that many readers probably understand much of what they are reading, but not every single word”, thus the need for a website offering free online bilingual translation dictionaries.

In 2010, WordReference has also offered a monolingual dictionary in English as well as dictionaries from English to other languages (Arabic, Chinese, Czech, Greek, Japanese, Korean, Polish, Portuguese, Romanian, Turkish), and vice versa. For the Spanish language, there was a monolingual dictionary, a dictionary of synonyms, a Spanish-French dictionary and a Spanish-Portuguese dictionary. Conjugation tables were available for French, Italian and Spanish. Monolingual dictionaries were available for German and Russian.

WordReference Mini was a miniature version of the site to be embedded into other sites, for example sites teaching languages online. A mobile device version was available for dictionaries from English to French, Italian and Spanish, and vice versa, with other language pairs to come.

As stated by Michael Kellogg: “Today [in 2011], I have three main goals with this website. First, continue to create free online bilingual dictionaries for English to many other languages. I strive to offer translations for "all" English words, terms, idioms, sayings, etc. Second, provide the world's best language forums; and third, continue to innovate to produce the best website and tools for the world.”

TRANSLATORS

[Summary] The internet became a mandatory tool for translators as “a vital and endless source of information”, as stated by Marcel Grangier, the head of the French Section of Central Linguistic Services, which means he was in charge of organizing translation matters into French for the linguistic services of the Swiss government. He explained in January 1999: “To work without the internet is simply impossible now. Apart from all the tools used (email, the electronic press, services for translators), the internet is for us a vital and endless source of information in what I'd call the 'non-structured sector' of the web. For example, when the answer to a translation problem can't be found on websites presenting information in an organized way, in most cases search engines allow us to find the missing link somewhere on the network.” His services also offered an online directory called “Dictionnaires Électroniques” (Electronic Dictionaries) with links to most quality dictionaries available for free on the web.

The internet became a mandatory tool for translators as “a vital and endless source of information”, as stated by Marcel Grangier, the head of the French Section of Central Linguistic Services.

Marcel Grangier was in charge of organizing translation matters into French for the linguistic services of the Swiss government. He explained in January 1999: “To work without the internet is simply impossible now. Apart from all the tools used (email, the electronic press, services for translators), the internet is for us a vital and endless source of information in what I'd call the 'non-structured sector' of the web. For example, when the answer to a translation problem can't be found on websites presenting information in an organized way, in most cases search engines allow us to find the missing link

somewhere on the network.

Our website was first conceived as an intranet service for translators in Switzerland, who often deal with the same kind of material as the Federal government's translators. Some parts of it are useful to any translators, wherever they are. The section "Dictionnaires Électroniques" [Electronic Dictionaries] is only one section of the website. Other sections deal with administration, law, the French language, and general information. The site also hosts the pages of the Conference of Translation Services of European States (COTSOES)."

"Dictionnaires Électroniques" is a extensive directory of free dictionaries available online, with five main sections: abbreviations and acronyms, monolingual dictionaries, bilingual dictionaries, multilingual dictionaries, and geographical information. The index could also be searched by keywords. It was later transferred on the new website of COTSOES.

According to Marcel Grangier, "we can see multilingualism on the internet as a happy and irreversible inevitability. So we have to laugh at the doomsayers who only complain about the supremacy of English. Such supremacy is not wrong in itself, because it is mainly based on statistics (more PCs per inhabitant, more people speaking English, etc.). The answer is not to 'fight' English, much less whine about it, but to build more sites in other languages. As a translation service, we also recommend that websites be multilingual. (...) The increasing number of languages on the internet is inevitable and can only boost multicultural exchanges. For this to happen in the best possible circumstances, we still need to develop tools to improve compatibility. Fully coping with accents and other characters is only one example of what can be done."

Maria Victoria Marinetti was a translator from French to Spanish living near Geneva, Switzerland, with a doctorate in engineering from Mexico. She

wrote in August 1999: "I have access to a large number of global information, which is very interesting for me. I can also regularly send or receive files back and forth. The internet allows be to receive or send general and technical translations from French into Spanish, and vice versa, and to correct texts in Spanish. In the technical or chemical fields, I offer a technical assistance, as well as information about exporting high-tech equipment to Mexico or to other Latin American countries."

As for multilingualism, "it is very important to be able to communicate in various languages. I would even say this is mandatory, because the information given on the internet is meant for the whole world, so why wouldn't we get this information in our language or in the language we wish? Worldwide information, but no broad choice for languages, this would be quite a contradiction, wouldn't it?"

In 2000, the internet was multilingual, with half of its users having a mother tongue other than English, but the language barrier was far from gone. If any language was now available on the web, many users were monolingual, and even plurilingual users couldn't read all languages. Bridges were needed between language communities to improve the flow of information in other languages, including by offering better translation software and by offering tools for all languages, and not only the dominant ones.

1999 > THE NEED FOR BILINGUAL INFORMATION ONLINE

[Summary] With the web spreading worldwide, bilingual information online became mandatory, as stated by Henk Slettenhaar, a professor in communication technologies at Webster University, Geneva, Switzerland,

and a trilingual European. Henk spent his childhood in Holland, has taught his courses in English and has lived in neighboring France. He wrote in August 1999: "There are two main categories of websites in my opinion. The first one is the global outreach for business and information. Here the language is definitely English first, with local versions where appropriate. The second one is local information of all kinds in the most remote places. If the information is meant for people of an ethnic and/or language group, it should be in that language first, with perhaps a summary in English. We have seen lately how important these local websites are — in Kosovo and Turkey, to mention just the most recent ones. People were able to get information about their relatives through these sites."

With the web spreading worldwide, bilingual information online became mandatory, as stated by Henk Slettenhaar, a professor in communication technologies at Webster University, Geneva, Switzerland, and a trilingual European.

Henk spent his childhood in Holland, has taught his courses in English and has lived in neighboring France. He wrote in December 1998: "I see multilingualism as a very important issue. Local communities that are on the web should principally use the local language for their information. If they want to present it to the world community as well, it should be in English too. I see a real need for bilingual websites. I am delighted there are so many offerings in the original language now. I much prefer to read the original with difficulty than getting a bad translation."

Henk added in August 1999: "There are two main categories of websites in my opinion. The first one is the global outreach for business and information. Here the language is definitely English first, with local versions where appropriate. The second one is local information of all kinds in the most

remote places. If the information is meant for people of an ethnic and/or language group, it should be in that language first, with perhaps a summary in English. We have seen lately how important these local websites are — in Kosovo and Turkey, to mention just the most recent ones. People were able to get information about their relatives through these sites."

Geoffrey Kingscott, managing director of Praetorius, a language consultancy in applied languages, wrote in September 1998: "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* [a magazine for linguists] 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 it 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."

Steven Krauwer, coordinator of ELSNET (European Network of

Excellence in Human Language Technologies), explained in September 1998: "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."

What practical solutions would he suggest? He answered in August 1999: "At the author end: better education of web authors to use combinations of modalities to make communication more effective across language barriers (and not just for cosmetic reasons). At the server end: more translation facilities à la AltaVista (quality not impressive, but always better than nothing). At the browser end: more integrated translation facilities (especially for the smaller languages), and more quick integrated dictionary lookup facilities."

Bruno Didier, webmaster of the Pasteur Institute's library, wrote in August 1999: "The internet doesn't belong to any one nation or language. It is a vehicle for culture, and the first vector of culture is language. The more languages there are on the net, the more cultures will be represented there. I don't think we should give in to the kneejerk temptation to translate webpages into a largely universal language. Cultural exchanges will only be real if we are prepared to meet with the other culture in a genuine way. And this effort involves understanding the other culture's language. This is very idealistic of course. In practice, when I am monitoring, I curse Norwegian or Brazilian websites where there isn't any English."

Alain Bron, a consultant in information systems and a writer, explained in January 1999: "Different languages will still be used for a long time to come

and this is healthy for the right to be different. The risk is of course an invasion of one language to the detriment of others, and with it the risk of cultural standardization. I think online services will gradually emerge to get around this problem. First, translators will be able to translate and comment on texts by request, but mainly sites with a large audience will provide different language versions, just as the audiovisual industry does now."

In spring 2000, non-English-speaking users reached 50%. 78% of webpages were still in English in September 2000.

2000 > ONLINE ENCYCLOPEDIAS AND DICTIONARIES

[Summary] The first reference encyclopedias and dictionaries available online stemmed from print versions. Britannica.com was available in December 1999 as the web version of the 32-volume Encyclopaedia Britannica, first for free and then for a fee. The French-language WebEncyclo from Editions Atlas was available at the same time, for free, as well as the Encyclopaedia Universalis, for a fee. The first major online dictionaries also stemmed from print versions, for example the free Merriam-Webster Online launched in 1996, that included the Webster Dictionary, the Webster Thesaurus, and other tools. The French-language "Dictionnaire Universel Francophone en Ligne " from Hachette was available for free in 1997. The online version of the 20-volume Oxford English Dictionary (OED) was available in March 2000 for a fee. Designed directly for the web, the Grand Dictionnaire Terminologique (GDT) was launched in September 2000 in Quebec as the largest free French-English terminology dictionary, and quickly praised by linguists worldwide.

The first reference encyclopedias and dictionaries available online stemmed from print versions.

Encyclopedias

Britannica.com was launched in December 1999 as the digital equivalent of the 32 volumes of the 15th edition of the Encyclopaedia Britannica. The website was available for free, as a complement to the print and CD-ROM versions for sale, with a selection of articles from 70 magazines, a guide to the best websites, a selection of books, etc., all searchable through a single search engine. In September 2000, the site was among the top 100 websites in the world. In July 2001, the website, not free anymore, could be searched for a monthly or annual fee. In 2009, Britannica.com opened its website to external contributors, with registration required to write and edit articles.

Launched by Editions Atlas in December 1999 and stemming from a print encyclopedia, Webencyclo was the first main French-language online encyclopedia available for free. It was searchable by keyword, topic and media (i.e. maps, links, photos, illustrations). A call for papers invited specialists in a given topic to become external contributors and submit articles in a section called "Webencyclo Contributif". Later on, a free registration was required to use the online encyclopedia.

Launched at the same time, the website of the print French-language Encyclopedia Universalis included 28,000 articles by 4,000 contributors, available for an annual subscription fee, with a number of articles available for free.

Dictionaries

Merriam-Webster, a well-known publisher of dictionaries, launched in 1996 the website "Merriam-Webster Online: The Language Center" to give

free access to online resources stemming from several print reference works: Webster Dictionary, Webster 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 website's goal was also to help track down definitions, spellings, pronunciations, synonyms, vocabulary exercises, and other key facts about words and language.

The "Dictionnaire Universel Francophone en Ligne" (Universal French-Language Online Dictionary) was the web version of the "Dictionnaire Universel Francophone", published by Hachette in partnership with AUPELF-UREF (which later became AUF: Agence Universitaire de la Francophonie - University Agency of Francophony). The dictionary included not only standard French but also the French-language words and expressions used worldwide. French was spoken by 500 million people in 50 countries. As a side remark, English and French are the only official and/or cultural languages widely spread on five continents.

The online version (for a subscription fee) of the 20-volume Oxford English Dictionary (OED) was launched in March 2000 by Oxford University Press (OUP), followed by a quarterly update with around 1,000 new or revised entries. Two years later, Oxford University Press launched Oxford Reference Online (ORO), a comprehensive encyclopedia designed directly for the web and also available for a subscription fee. Its 60,000 webpages and one million entries could represent the equivalent of 100 print encyclopedias.

The GDT from Quebec

With 3 million terms related to industry, science and commerce, the GDT (Grand Dictionnaire Terminologique - Main Terminological Dictionary) was the largest French-English online terminology dictionary. The GDT was

designed directly for the web by OQLF (Office Québécois de la Langue Française - Quebecois Office of the French Language) and launched in September 2000 as a free service. The GDT was a technological challenge, and the result of a partnership between OQLF, author of the dictionary, and Semantix, a company specialized in linguistic software. The GDT had 1.3 million individual visits during the first month, with peaks of 60,000 visits per day, which certainly contributed to better translations. The database was then maintained by Convera Canada, with 3.5 million visits per month in February 2003. A revamped version of the GDT went online in March 2003, with the database maintained by OQLF itself, and the addition of Latin as a third language.

2000 > THE WEB PORTAL YOURDICTIONARY.COM

[Summary] Robert Beard, a language teacher at Bucknell University, in Lewisburg, Pennsylvania (USA), co-founded yourDictionary.com in February 2000 as a follow-up of his first website, A Web of Online Dictionaries (included in the new one), launched in 1995 as a directory of online dictionaries (with 800 links in fall 1998) and other linguistic resources such as thesauri, vocabularies, glossaries, grammars and language textbooks. yourDictionary.com included 1,800 dictionaries in 250 languages in September 2003, and 2,500 dictionaries in 300 languages in April 2007. As a portal for all languages without any exception, the site also offered a section for endangered languages, called the Endangered Language Repository.

Five years before co-founding yourDictionary.com in February 2000, as the portal for all languages without any exception, Robert Beard created the

website A Web of Online Dictionaries (WOD) in 1995.

Robert Beard was a language teacher at Bucknell University, in Lewisburg, Pennsylvania (USA). In September 1998, his website provided an index of 800 online dictionaries in 150 languages, as well as sections for multilingual dictionaries, specialized English dictionaries, thesauri and other vocabulary aids, language identifiers and guessers, an index of dictionary indices, the “Web of Online Grammars”, and the “Web of Linguistic Fun”, i.e. linguistics for non-specialists.

Robert Beard wrote in September 1998: "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."

Fifteen months later, Robert Beard included his website into a larger project, yourDictionary.com, that he co-founded in early 2000.

He wrote in January 2000: "The new website is an index of 1,200+ dictionaries in more than 200 languages. Besides the WOD, the new website includes a word-of-the-day-feature, word games, a language chat room, the old 'Web of Online Grammars' (now expanded to include additional language resources), the 'Web of Linguistic Fun', multilingual dictionaries; specialized English dictionaries; thesauri and other vocabulary aids; language identifiers and guessers, and other features; dictionary indices. yourDictionary.com will hopefully be the premiere language portal and the largest language resource

site on the web. It is now actively acquiring dictionaries and grammars of all languages with a particular focus on endangered languages. It is overseen by a blue ribbon panel of linguistic experts from all over the world. (...)

Indeed, yourDictionary.com has lots of new ideas. We plan to work with the Endangered Language Fund in the U.S. and Britain to raise money for the Foundation's work and publish the results on our site. We will have language chat rooms and bulletin boards. There will be language games designed to entertain and teach fundamentals of linguistics. The Linguistic Fun page will become an online journal for short, interesting, yes, even entertaining, pieces on language that are based on sound linguistics by experts from all over the world."

As the portal for all languages without any exception, yourDictionary.com offered a section for endangered languages called the Endangered Language Repository.

As explained by Robert Beard: "Languages that are endangered are primarily languages without writing systems at all (only 1/3 of the world's 6,000+ languages have writing systems). I still do not see the web contributing to the loss of language identity and still suspect it may, in the long run, contribute to strengthening it. More and more Native Americans, for example, are contacting linguists, asking them to write grammars of their language and help them put up dictionaries. For these people, the web is an affordable boon for cultural expression."

How about the future of the web? "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."

2000 > PROJECT GUTENBERG AND LANGUAGES

[Summary] Project Gutenberg is a visionary project launched by Michael Hart in July 1971 to create free electronic versions of literary works and disseminate them worldwide. In 2010, Project Gutenberg offered more than 33,000 high-quality ebooks being downloaded by the tens of thousands every day, and websites in the United States, Australia, Europe and Canada, with 40 mirror sites worldwide. Project Gutenberg mainly offers ebooks in English, but multilingualism has been one of its priorities since the late 1990s. French is the second language of the project. There were ebooks in 60 languages in December 2010, thanks to the patient work of Distributed Proofreaders, a website created in 2000 to share the proofreading of ebooks between hundreds of volunteers in many countries.

Project Gutenberg is a visionary project launched in July 1971 by Michael Hart to create free electronic versions of literary works and disseminate them worldwide. In the 15th century, Gutenberg allowed anyone to have print books for a small cost. In the 21st century, Project Gutenberg would allow anyone to have a digital library at no cost.

Michael worked from Illinois, typing in books from public domain, for example the Bible and the complete works of Shakespeare, first alone, then with the help of a few volunteers.

His project got a major boost with the invention of the web in 1990. 95% of internet users were native English speakers in the mid-1990s, so most books were in English.

Project Gutenberg was also inspiring other digital libraries in Europe. Projekt Runeberg was launched in Sweden in 1994 to digitize Nordic (Scandinavian) literature from public domain. Projekt Gutenberg-DE was launched in Germany in 1994 to digitize German literature from public domain.

French was the second language of Project Gutenberg, and still is now. The first ebooks released in French were six works by Stendhal and two works by Jules Verne, all released in early 1997. Three novels by Jules Verne were already available in English in 1994. Since then, Jules Verne has always stayed on the top list of the most downloaded authors.

In October 1997, Michael Hart wrote about producing more works in other languages than English in the Project Gutenberg newsletter. In early 1998, on top of ten French ebooks, there were a few ebooks in German, Italian, Spanish and Latin. Released in May 1999, eBook #2000 was “Don Quijote” (1605), by Cervantes, in Spanish, its original language. In July 1999, Michael wrote in an email interview: "I am publishing in one new language per month right now, and will continue as long as possible."

The project got a new boost with the launching of Distributed Proofreaders, a website created in October 2000 by Charles Franks to share the proofreading of ebooks between hundreds of volunteers living in many countries.

Released in April 2002, eBook #5000 was “The Notebooks of Leonardo da Vinci” (written in the early 16th century), as an English translation from Italian, its original language. Since its release, it has regularly stayed in the top 100 downloaded ebooks.

There were works in 25 languages in early 2004, in 42 languages in July 2005, including Sanskrit and the Mayan languages, and in 59 languages in

October 2010. The ten main languages were English (with 28,441 ebooks on 7 October 2010), French (1,659 ebooks), German (709 ebooks), Finnish (536 ebooks), Dutch (496 ebooks), Portuguese (473 ebooks), Chinese (405 ebooks), Spanish (295 ebooks), Italian (250 ebooks), and Greek (101 ebooks). The next languages were Latin, Esperanto, Swedish and Tagalog.

When machine translation will be judged 99% satisfactory, we may be able to read literary classics in a choice of many languages. The machine translated ebooks won't compete with the work of literary translators and their labor of love during days and months if not years, but they will allow readers to get the gist of some literary works that have never been translated so far, or only translated in a few languages for commercial reasons.

The output of translation software could then be proofread by human translators, in a similar way the output of OCR software is proofread by the volunteers of Distributed Proofreaders. So, may be, we will see the creation of Distributed Translators one day, as a partner or sister project of Distributed Proofreaders and Project Gutenberg.

2001 > WIKIPEDIA, A COLLABORATIVE ENCYCLOPEDIA

[Summary] Wikipedia was launched in January 2001 by Jimmy Wales and Larry Sanger (Larry resigned later on) as a global free collaborative online encyclopedia, financed by donations, with no advertising. Its website is a wiki, which means that anyone can write, edit, correct and improve information throughout the encyclopedia, with people contributing under a pseudonym. The articles stay the property of their authors, and can be freely used according to Creative Commons or GFDL (GNU Free Documentation License). Wikipedia quickly became the largest reference website. It was in

the top ten websites in December 2006, and in the top five websites in 2008. In May 2007, Wikipedia had 7 million articles in 192 languages, including 1.8 million articles in English, 589,000 articles in German, 500,000 articles in French, 260,000 articles in Portuguese, and 236,000 articles in Spanish. Wikipedia celebrated its tenth anniversary in January 2011 with 17 million articles in 270 languages et 400 million individual visits per month for all websites.

Wikipedia was launched in January 2001 by Jimmy Wales and Larry Sanger (Larry resigned later on) as a global free collaborative online encyclopedia.

Wikipedia was financed by donations, with no advertising. Its website is a wiki, which means that anyone can write, edit, correct and improve information throughout the encyclopedia, with people contributing under a pseudonym. The articles stay the property of their authors, and can be freely used according to Creative Commons or GFDL (GNU Free Documentation License).

Wikipedia is hosted by the Wikimedia Foundation, founded in June 2003, which has run a number of other projects, beginning with Wiktionary (launched in December 2002) and Wikibooks (launched in June 2003), followed by Wikiquote, Wikisource (texts from public domain), Wikimedia Commons (multimedia), Wikispecies (animals and plants), Wikinews and Wikiversity (textbooks).

Wikipedia quickly became the largest reference website, with thousands of people contributing worldwide. In December 2004, Wikipedia had 1.3 million articles by 13,000 contributors in 100 languages. In December 2006, Wikipedia was among the top ten sites on the web, with 6 million articles. In

May 2007, Wikipedia had 7 million articles in 192 languages, including 1.8 million articles in English, 589,000 articles in German, 500,000 articles in French, 260,000 articles in Portuguese, and 236,000 articles in Spanish. In 2008, Wikipedia was in the top five websites. In September 2010, Wikipedia had 14 million articles in 272 languages, including 3.4 million articles in English, 1.1 million articles in German and 1 million articles in French. Wikipedia celebrated its tenth anniversary in January 2011 with 17 million articles in 270 languages et 400 million individual visits per month for all websites.

Wikipedia also inspired many other projects over the years, for example Citizendium, launched in 2007 as a pilot project to build a new encyclopedia.

Citizendium, an acronym for “The Citizen’s Compendium”, was launched in March 2007 at the initiative of Larry Sanger, who co- founded Wikipedia with Jimmy Wales in January 2001, but resigned later on over policy and content quality issues, as well as the use of anonymous pseudonyms.

Citizendium is a wiki project open to public collaboration, but combining "public participation with gentle expert guidance". The project is experts-led, not experts-only. Contributors use their own names, and they are guided by expert editors. As explained by Larry in his essay "Toward a New Compendium of Knowledge", posted in September 2006 and updated in March 2007: "Editors will be able to make content decisions in their areas of specialization, but otherwise working shoulder-to-shoulder with ordinary authors." There are also constables who make sure the rules are respected.

There were 1,100 high-quality articles, 820 authors, and 180 editors in March 2007, 11,800 articles in August 2009, and 15,000 articles in September 2010. Citizendium wants to act as a prototype for upcoming large scale knowledge-building projects that would deliver reliable reference, scholarly and educational content.

2001 > UNL, A DIGITAL METALANGUAGE PROJECT

[Summary] The UNDL Foundation (UNDL: Universal Networking Digital Language) was founded in January 2001 to develop and promote the UNL (Universal Networking Language) project. The UNL project was launched in 1996 as a major digital metalanguage project by the Institute of Advanced Studies (IAS) of the United Nations University (UNU) in Tokyo, Japan. As explained in 1998 on the bilingual English-Japanese website: "UNL 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 web browsers, on the internet, and will be compatible with standard network servers." At the time, 120 researchers worldwide were working on a multilingual project in 16 languages (Arabic, Brazilian, Chinese, English, French, German, Hindi, Indonesian, Italian, Japanese, Latvian, Mongolian, Russian, Spanish, Swahili, Thai).

The UNDL Foundation (UNDL: Universal Networking Digital Language) was founded in January 2001 to develop and promote the UNL (Universal Networking Language) project.

The UNL project was launched in 1996 as a major digital metalanguage project by the Institute of Advanced Studies (IAS) of the United Nations University (UNU) in Tokyo, Japan.

As explained in 1998 on the bilingual English-Japanese website: "UNL 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 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."

At the time, 120 researchers worldwide were working on a multilingual project in 16 languages (Arabic, Brazilian, Chinese, English, French, German, Hindi, Indonesian, Italian, Japanese, Latvian, Mongolian, Russian, Spanish, Swahiki, Thai). After things worked with 16 languages, other UN languages would be included in 2000.

UNL was meant to become the HTML of linguistic content. Possible applications would be multilingual email, multilingual information, active dictionaries for reading foreign languages online, and machine translation for navigating the web and monitoring websites.

The project was also important from a political and cultural point of view, as the first project building up tools for all languages on the internet, i.e. main languages as well as minority languages.

The UNDL Foundation (UNDL: Universal Networking Digital Language) was founded in January 2001 to develop and promote the UNL project, and became a partner of the United Nations.

The definition of UNL has evolved over the years. According to the UNDL Foundation's website in 2010: "UNL is a computer language that enables computers to process information and knowledge. It is designed to replicate the functions of natural languages. Using UNL, people can describe all

information and knowledge conveyed by natural languages for computers. As a result, computers can intercommunicate through UNL and process information and knowledge using UNL, thus providing people with a Linguistic Infrastructure (LI) in computers and the internet for distributing, receiving and understanding multilingual information. Such multilingual information can be accessed by natural languages through the UNL System. UNL, as a language for expressing information and knowledge described in natural languages, has all the components corresponding to that of a natural language.”

2001 > A MARKET FOR LANGUAGE TRANSLATION SOFTWARE

[Summary] The development of electronic commerce boosted language translation software, products and services targeting the general public, language professionals, and companies localizing their websites. The software, products and services were developed for example by Alis Technologies, Globalink, Lernout & Hauspie, Softissimo and IBM. In March 2001, IBM embarked on a growing translation market with a high-end professional product, the WebSphere Translation Server. The software could instantly translate webpages, emails and chats from/into several languages (Chinese, English, French, German, Italian, Japanese, Korean, Spanish). It could process 500 words per second and add terminology to the software. Computer-assisted translation (CAT) software were developed for professional translators, based on “translation memory” with terminology processing in real time, for example Wordfast, created in 1999 by Yves Champollion. Wordfast could be used on any platform (Windows, Mac, Linux), and was compatible with the software of other key players like IBM and SDL Trados.

The development of electronic commerce boosted language translation software, products and services targeting the general public, language professionals, and companies localizing their websites.

The software, products and services were developed for example by Alis Technologies, Globalink, Lernout & Hauspie, Softissimo and IBM.

In March 2001, IBM embarked on a growing translation market with a high-end professional product, the WebSphere Translation Server. The software could instantly translate webpages, emails and chats from/into several languages (Chinese, English, French, German, Italian, Japanese, Korean, Spanish). It could process 500 words per second and add terminology to the software.

Machine translation can be defined as the automated process of translating a text from one language to another language. MT analyzes the text in the source language and automatically generates the corresponding text in the target language. With the lack of human intervention during the translation process, machine translation differs from computer-assisted translation (CAT), based on interaction between the translator and the computer.

Computer-assisted translation (CAT) software were developed for professional translators, based on “translation memory” with terminology processing in real time, for example Wordfast, created in 1999 by Yves Champollion. Wordfast was compatible with the software of other key players like IBM and SDL Trados. Available for any platform (Windows, Mac, Linux), Wordfast had 14,000 customers worldwide in 2010, including the United Nations, Coca- Cola and Sony.

According to Tim McKenna, a writer and philosopher interviewed in October 2000: "When software gets good enough for people to chat or talk on the web in real time in different languages, then we will see a whole new world appear before us. Scientists, political activists, businesses and many more groups will be able to communicate immediately without having to go through mediators or translators."

A further step could be "transcultural, transnational transparency", as stated in September 1998 by Randy Hobler, a consultant in internet marketing of translation software and services: "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. (...)

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. 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."

2004 > THE WEB 2.0, COMMUNITY AND SHARING

[Summary] The term "web 2.0" was invented in 2004 by Tim O'Reilly, founder of O'Reilly Media, a publisher of computer books, as a title for a series of conferences he was organizing. The web 2.0 was based on community and sharing, with a wealth of websites whose content was supplied by users, such as blogs, wikis, social networks and collaborative encyclopedias. Wikipedia, Facebook and Twitter, of course, but also tens of thousands of others. The web 2.0 may begin to fulfill the dream of Tim Berners-Lee, who invented the web in 1990, and wrote in an essay dated April 1998: "The dream behind the web is of a common information space in which we communicate by sharing information. Its universality is essential: the fact that a hypertext link can point to anything, be it personal, local or global, be it draft or highly polished. ("The World Wide Web: A very short personal history", available on his webpage on the W3C website)

The term "web 2.0" was invented in 2004 by Tim O'Reilly, founder of O'Reilly Media, a publisher of computer books, as a title for a series of conferences he was organizing.

The web 2.0 was based on community and sharing, with a wealth of websites whose content was supplied by users, such as blogs, wikis, social networks and collaborative encyclopedias. Wikipedia, Facebook and Twitter, of course, but also tens of thousands of others.

The web 2.0 may begin to fulfill the dream of Tim Berners-Lee, who invented the web in 1990, and wrote in an essay dated April 1998: "The dream behind the web is of a common information space in which we communicate by sharing information. Its universality is essential: the fact that

a hypertext link can point to anything, be it personal, local or global, be it draft or highly polished.” (“The World Wide Web: A very short personal history”, available on his webpage on the W3C website)

The first blog was launched in 1997. A blog is an online diary kept by a person or a group, usually in reverse chronological order, and can be updated every minute or once a month. There were 14 million blogs worldwide in July 2005, with 80,000 new blogs per day. According to Technorati, the first blog search engine, there were 65 million blogs in December 2006, with 175,000 new blogs per day. Some blogs are devoted to photos (photoblogs), music (audioblogs or podcasts), and videos (vlogs or videoblogs).

The wiki concept became quite popular in 2000. Deriving from the Hawaiian term "wiki" ("fast"), a wiki is a website allowing multiple users to collaborate online on the same project. Users can contribute to drafting content, editing it, improving it, and updating it. The software can be simple or more elaborate. A simple program handles text and hyperlinks. With a more elaborate program, one can embed images, charts, tables, etc. The most famous wiki is Wikipedia.

Facebook was founded in February 2004 by Mark Zuckerberg and his fellow students as a social network. Originally created for the students of Harvard University, it was then available to students from any university in the U.S. before being open to anyone worldwide in September 2006, to connect with relatives, friends and strangers. Facebook was the second most visited website after Google, with 500 million users in June 2010, while sparking debates on privacy issues.

Founded in 2006 by Jack Dorsey and Biz Stone, Twitter is a social networking and micro-blogging tool to send free short messages of 140 characters maximum, called tweets, via the internet, IM or SMS. Sometimes described as the SMS of the internet, Twitter gained worldwide popularity,

with 106 million users in April 2010, and 300,000 new users per day. As for tweets, there were 5,000 per day in 2007, 300,000 in 2008, 2.5 million in 2009, 50 million in January 2010, and 55 million in April 2010, with the archiving of public tweets by the Library of Congress as a reflection of the trends of our time.

We now try to fulfill the second part of Tim Berners-Lee's dream, according to his essay dated April 1998: "There was a second part of the dream, too, dependent on the web being so generally used that it became a realistic mirror (or in fact the primary embodiment) of the ways in which we work and play and socialize. That was that once the state of our interactions was online, we could then use computers to help us analyze it, make sense of what we are doing, where we individually fit in, and how we can better work together."

2007 > THE ISO 639-3 STANDARD TO IDENTIFY LANGUAGES

[Summary] The first standard to identify languages was ISO 639-1, adopted by the International Organization for Standardization (ISO) in 1988 as a set of two-letter identifiers. The ISO 639-2 standard followed in 1998 as a set of three-letter codes identifying 400 languages. Published by SIL International, the Ethnologue, an encyclopedic catalog of living languages, had also developed its own three-letter codes in its database since 1971, with their inclusion in the publication itself since 1984 (10th edition). ISO 639-2 quickly became outdated. In 2002, at the invitation of the International Organization for Standardization, SIL International prepared a new standard that reconciled the complete set of identifiers used in the Ethnologue with the identifiers already in use in ISO 639-2, as well as identifiers developed by the Linguist List to handle ancient and constructed languages. Published in 2007,

the ISO 639-3 standard provided three-letter codes for identifying 7,589 languages. SIL International was named the registration authority for the inventory of language identifiers.

Published in 2007, the ISO 639-3 standard provided three-letter codes for identifying 7,589 languages.

The first standard to identify languages was ISO 639-1, adopted by the International Organization for Standardization (ISO) in 1988 as a set of two-letter language identifiers.

The ISO 639-2 standard followed in 1998 as a set of three-letter codes identifying 400 languages. The standard was a convergence of ISO 639-1 and the ANSI Z39.53 standard (ANSI: American National Standards Institute). The ANSI standard corresponded to the MARC (Machine Readable Cataloging) language codes, a set of three-letter identifiers developed by the library community and adopted as an American National Standard in 1987.

Published by SIL International, the Ethnologue, an encyclopedic catalog of living languages, had also developed its own three-letter codes in its database since 1971, with the inclusion in the encyclopedia itself from the 10th edition (1984) onwards.

ISO 639-2 quickly became insufficient because of the small number of languages it could handle. In 2002, at the invitation of the International Organization for Standardization, SIL International prepared a new standard that reconciled the complete set of codes used in the Ethnologue with the codes already in use in ISO 639-2, as well as codes developed by the Linguist List — a main distribution list for linguists — to handle ancient and constructed languages.

Approved in 2006 and published in 2007, the ISO 639-3 standard provided three-letter codes for identifying 7,589 languages, with a list of languages as complete as possible, living and extinct, ancient and reconstructed, major and minor, and written and unwritten. SIL International was named the registration authority for the inventory of language identifiers, and administers the annual cycle for changes and updates.

2007 > GOOGLE TRANSLATE

[Summary] Launched by Google in October 2007, Google Translate is a free online language translation service that instantly translates a section of text, document or webpage into another language. Users paste texts in the web interface or supply an hyperlink. The automatic translations are produced by statistical analysis rather than traditional rule-based analysis. Prior to this date, Google used a Systran based translator like Babel Fish in Yahoo! As an automatic translation tool, Google Translate can help the reader understand the general content of a foreign language text, but doesn't deliver accurate translations. In 2009, the text could be read by a speech program, with new languages added over the months. Released in June 2009, Google Translator Toolkit is a web service allowing (human) translators to edit the translations automatically generated by Google Translate. In January 2011, people could choose different translations for a word in Google Translate.

Launched by Google in October 2007, Google Translate is a free online language translation service that instantly translates a section of text, document or webpage into another language.

Users paste texts in the web interface or supply an hyperlink. The automatic translations are produced by statistical analysis rather than traditional rule-based analysis.

As an automatic translation tool, Google Translate can help the reader understand the general content of a foreign language text, but doesn't deliver accurate translations.

Prior to this date, Google used a Systran based translator like Babel Fish in Yahoo!, with several stages for the language options:

First stage: English to French, German, and Spanish, and vice versa.

Second stage: English to Portuguese and Dutch, and vice versa.

Third stage: English to Italian, and vice versa.

Fourth stage: English to simplified Chinese, Japanese and Korean, and vice versa.

Fifth stage (April 2006): English to Arabic, and vice versa.

Sixth stage (December 2006): English to Russian, and vice versa.

Seventh stage (February 2007): English to traditional Chinese, and simplified Chinese to traditional Chinese, and vice versa.

Here were the first language options for Google's translation system:

First stage (October 2007): All language pairs previously available were available in any language combination.

Second stage: English to Hindi, and vice versa.

Third stage (May 2008): Bulgarian, Croatian, Czech, Danish, Finnish, Greek, Norwegian, Polish, Romanian, Swedish, with any combination.

Fourth stage (September 2008): Catalan, Filipino, Hebrew, Indonesian, Latvian, Lithuanian, Serbian, Slovak, Slovene,

Ukrainian, Vietnamese.

Fifth stage (January 2009): Albanian, Estonian, Galician, Hungarian, Maltese, Thai, Turkish.

Sixth stage (June 2009): Persian.

Seventh stage (August 2009): Afrikaans, Belarussian, Icelandic, Irish, Macedonian, Malay, Swahili, Welsh, Yiddish.

Eighth stage (January 2010): Haitian Creole.

Ninth stage (May 2010): Armenian, Azeri, Basque, Georgian, Urdu.

Tenth stage (October 2010): Latin.

Etc.

A speech program was launched in 2009 to read the translated text, with new languages added over the months. In January 2011, people could choose different translations for a word in Google Translate.

Google Translator Toolkit is a web service allowing (human) translators to edit the translations automatically generated by Google Translate. Translators can also use shared translations, glossaries and translation memories. Starting in June 2009 with English as a source language and 47 target languages, Google Translator Toolkit supported 100,000 language pairs in May 2011, with 345 source languages into 345 target languages.

2009 > 6,909 LIVING LANGUAGES IN THE ETHNOLOGUE

[Summary] 6,909 living languages were cataloged in the 16th edition (2009) of “The Ethnologue: Languages of the World”, an encyclopedic reference work freely available on the web since 1996, with a print book for sale. As stated by Barbara Grimes, its editor from 1971 to 2000, the Ethnologue is “a catalog of the languages of the world, with information

about where they are spoken, an estimate of the number of speakers, what language family they are in, alternate names, names of dialects, other socio-linguistic and demographic information, dates of published Bibles, a name index, a language family index, and language maps." A core team of researchers in Dallas, Texas, has been helped by thousands of linguists gathering and checking information worldwide. A new edition of the Ethnologue is published approximately every four years.

6,909 living languages were cataloged in the 16th edition (2009) of "The Ethnologue: Languages of the World", an encyclopedic reference work freely available on the web since 1996, with a print book for sale.

As stated by Barbara Grimes, its editor from 1971 to 2000, the Ethnologue is "a catalog of the languages of the world, with information about where they are spoken, an estimate of the number of speakers, what language family they are in, alternate names, names of dialects, other socio-linguistic and demographic information, dates of published Bibles, a name index, a language family index, and language maps."

A core team of researchers in Dallas, Texas, has been helped by thousands of linguists gathering and checking information worldwide. A new edition of the Ethnologue is published approximately every four years.

The Ethnologue has been an active research project since 1950. It was founded by Richard Pittman as a catalog of minority languages, to share information on language development needs around the world with his colleagues at SIL International and other language researchers.

Richard Pittman was the editor of the 1st to 7th editions (1951- 1969).

Barbara Grimes was the editor of the 8th to 14th editions (1971- 2000). In 1971, information was expanded from primarily minority languages to encompass all known languages of the world. Between 1967 and 1973, Barbara completed an in-depth revision of the information on Africa, the Americas, the Pacific, and a few countries of Asia. During her years as editor, the number of identified languages grew from 4,493 to 6,809. The information recorded on each language expanded so that the published work more than tripled in size.

In 2000, Raymond Gordon Jr. became the third editor of the *Ethnologue* and produced the 15th edition (2005).

In 2005, Paul Lewis became the editor, responsible for general oversight and research policy, with Conrad Hurd as managing editor, responsible for operations and database management, and Raymond Gordon as senior research editor, leading a team of regional and language-family focused research editors.

In the Introduction of the 15th edition (2009), the *Ethnologue* defines a language as such: "How one chooses to define a language depends on the purposes one has in identifying that language as distinct from another. Some base their definition on purely linguistic grounds. Others recognize that social, cultural, or political factors must also be taken into account. In addition, speakers themselves often have their own perspectives on what makes a particular language uniquely theirs. Those are frequently related to issues of heritage and identity much more than to the linguistic features of the language(s) in question."

As explained in the introduction, one feature of the database since its inception in 1971 has been a system of three-letter language identifiers (for example "fra" for French), that were included in the publication itself from the 10th edition (1984) onwards.

At the invitation of the International Organization for Standardization (ISO) in 2002, SIL International prepared a new standard that reconciled the complete set of codes used in the Ethnologue with the codes already in use in the ISO 639-2 standard (1998), that identified only 400 languages, as well as codes developed by Linguist List to handle ancient and constructed languages. Published in 2007, the ISO 639-3 standard provided three-letter codes for identifying nearly 7,500 languages. SIL International was named the registration authority for the inventory of language identifiers, and administers the annual cycle for changes and updates.

2010 > A UNESCO ATLAS FOR ENDANGERED LANGUAGES

[Summary] In 2010, UNESCO (United Nations Educational, Scientific and Cultural Organization) launched a free Interactive Atlas of the World's Languages in Danger. The online edition is a complement of the print edition (3rd edition, 2010), edited by Christopher Moseley, and available in English, French and Spanish, with previous editions in 1996 and 2001. 2,473 languages were listed on 4 June 2011, with a search engine by country and area, language name, number of speakers from/to, vitality and ISO 639-3 code. The language names have been indicated in English, French and Spanish transcriptions. Alternate names (spelling variants, dialects or names in non-Roman scripts) are also provided.

In 2010, UNESCO (United Nations Educational, Scientific and Cultural Organization) launched a free Interactive Atlas of the World's Languages in Danger.

The online edition is a complement of the print edition (3rd edition, 2010), edited by Christopher Moseley, and available in English, French and Spanish, with previous editions in 1996 and 2001.

2,473 languages were listed on 4 June 2011, with a search engine by country and area, language name, number of speakers from/to, vitality and ISO 639-3 code.

The language names have been indicated in English, French and Spanish transcriptions. Alternate names (spelling variants, dialects or names in non-Roman scripts) are also provided.

About language vitality

UNESCO's Language Vitality and Endangerment framework has established six degrees of vitality/endangerment: safe, vulnerable, definitely endangered, severely endangered, critically endangered, extinct.

“Safe” — not included in the atlas — means that the language is spoken by all generations and that intergenerational transmission is uninterrupted.

“Vulnerable” means that most children speak the language, but it may be restricted to certain domains, for example at home.

“Definitely endangered” means that children no longer learn the language as a mother tongue in the home.

“Severely endangered” means that the language is spoken by grandparents and older generations. While the parent generation may understand it, they don't speak it to children or among themselves.

“Critically endangered” means that the youngest speakers are grandparents and older, and they speak the language partially and infrequently.

“Extinct” means there are no speakers left. The atlas includes presumably extinct languages since the 1950s.

How to define an endangered language

When exactly is a language considered as endangered? As explained by UNESCO on the interactive atlas’ website: “A language is endangered when its speakers cease to use it, use it in fewer and fewer domains, use fewer of its registers and speaking styles, and/or stop passing it on to the next generation. No single factor determines whether a language is endangered.”

UNESCO experts have identified nine factors that should be considered together: (1) intergenerational language transmission; (2) absolute number of speakers; (3) proportion of speakers within the total population; (4) shifts in domains of language use; (5) response to new domains and media; (6) availability of materials for language education and literacy; (7) governmental and institutional language attitudes and policies including official status and use; (8) community members’ attitudes towards their own language; (9) amount and quality of documentation.

What are the causes of language endangerment and disappearance? “A language disappears when its speakers disappear or when they shift to speaking another language — most often, a larger language used by a more powerful group. Languages are threatened by external forces such as military, economic, religious, cultural or educational subjugation, or by internal forces such as a community’s negative attitude towards its own language. Today, increased migration and rapid urbanization often bring along the loss of traditional ways of life and a strong pressure to speak a dominant language that is — or is perceived to be — necessary for full civic participation and economic advancement.”

Copyright © 2012 Marie Lebert

*** END OF THE PROJECT GUTENBERG EBOOK THE WEB, A
MULTILINGUAL ENCYCLOPEDIA ***

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.

Table of Contents

[THE WEB, A MULTILINGUAL ENCYCLOPEDIA](#)
[TABLE OF CONTENTS](#)
[INTRODUCTION](#)
[THE FULL PROJECT GUTENBERG LICENSE](#)