1.1. The Internet - an overview

The Internet was started by the United States Department of Defence in the 60's. By the year 1991, the Internet has grown to include some 5,000 networks in over three dozen countries, serving over 700,000 host computers used by more than 4 million people (Leiner, 2001). In 1999, the global online population was 171 million web users with over 50% of them originating from outside the United States. By the year 2005, web users from outside the United States are expected to comprise 700 million of a total of one billion users (http://www.euromktg.com/globstats).

From the technical point of view, operating systems and software on the Internet are designed to support single-byte character sets such as ASCII (American Standard Code for Information Interchange). In this standard, letters are represented as digital or binary data and all letters are given a 'code number'. A single byte containing 7 binary digits represents each character. The Roman alphabet used in English Language and other European languages are easily accomodated by these software. Non-Roman scripts like Arabic and Hebrew need multi-byte character sets. Japanese and Chinese languages comprise thousands of characters; which is much more than what a single byte model can support. Even for languages with Roman script like French and German, full support for phonological features are not provided for. Most browsers are still unable to handle multilingual data presentation (Leiner, 2001). Due to the existence of such operating systems and software having a bias towards

supporting the roman alphabets, English has inevitably dominated the Internet.

1.1.1 Texts on the Internet

Texts on the Internet or technology texts are similar to written texts. However, technology texts are impermanent and are easily altered at any time. They can be copied, pasted and combined, thus allowing the creation of very large linked systems. These texts can be brought together according to the topic or subject and given a tag. These tagged texts allows complex searches of the linked information attached to it. Anyone using the Internet would understand that these tags link to more information.

Texts on the Internet also represent spoken language. There are templates and virtual identities on the Internet to simulate human interaction which approximates the norms of human communication. This simulates real world communication without being anchored in the constraints of physical contact. These properties formulated by Shortis (2001) will assist in the analysis of this study.

1.2. Languages on the Internet

Language use is usually associated with geographic boundaries. The Internet, however, has no geographic boundaries. It is global and borderless. Out of 462.02 million internet users in year 2002, 247.32

million were English speaking and 214.7 million were non-native speakers of English. In March 2003, 64.8% of the world online population were non-native speakers of English and only 36.2% were native speakers of English This does not mean that non-native speakers of English do not use English on the Internet. People in non-native countries like India, Pakistan and Phillipines use English to surf the Internet (http://www.euromktg.com/globstats).

Similarly, many of the people who access the Internet in other languages live in the United States and Canada.. They consist of Spanish, Chinese, Czech, French, Germans, Vietnamese and Korean descent (http://www.euromktg.com/globstats). Many Americans access the web in two languages, since for at least 45 million Americans , English is not the family language and 15 million do not speak English. For example, the American Civil Liberties Union of Florida, which has its States Office in a predominantly Spanish speaking Miami, has added a Spanish language page into its website. There is also a website in the Creole language to meet the needs of the country's large Haitian population (World, 1999).

Other languages on the Internet include the Russian language. A Director of a Russian Internet service once described the web as 'the ultimate act of intellectual colonialism' and thus, over 3000 Russian language media have been created in the Internet. In the year 2000, about 3,800 new Internet media have appeared in the Russian language (IATR/TASS, 2001).

In Asia, the majority of Internet users are the Chinese (90 million), the Japanese (69.7 million), the Koreans (29.2 million), the Malay speakers (8.7 million) and the Arabs (8.7 million) (http://www.euromktg.com/globstats). In China, China.com and Netease.com provides news in Chinese language. For example, www.gove.cn is a Chinese language site with no English index (China, 1999). The proliferation of Arabic language sites too has accelerated on the Internet. Emails and chat rooms are particularly popular in the Arab world, while online shopping and fast food services are also taking off. For instance, Ctlob.com provides delivery services to Egyptian online shoppers while Ajeeb.com provides instant translation in Arab language in addition to news and culture (Arabic, 2000).

Table 1.1

Chart of Web Content by Language in the Year 2000

Language	**%**
English	68.4
Japanese	5.9
German	5.8
Chinese	3.9
French	3.0
Spanish	2.4
Russian	1.9
Italian	1.6
Portuguese	1.4
Korean	1.3
Other	4.6
Total	313 billion web pages

Source : http://www.euromktg.com/globstats

Table 1.1 shows that other language websites are slowly proliferating on the Internet. In fact, with the growing number of non-native speakers of English on the Internet, there have been evidence that the demand for multilingual sites is growing steadily (Harvey, 2003).

However, creating sites in local languages other than English can be time consuming and costly. Although translation is one of the ways to access sites in local languages, translation services at the moment are still machine-made and the standard of translation is still questionable. Furthermore, apart from content translation, the tone, style and accent in these local languages have to be considered (Harvey, 2003). All these factors have made it more feasible and easier for English to dominate the Internet.

1.2.1. English on the Internet

Salverda (2002) describes English as the language of globalisation and modernisation. It allows communication between nations and people from different countries and is necessary and useful as a lingua franca.

English has only 256 characters which can easily be supported by most operating systems. Since the Internet started in the United States, naturally the language used is English. In a survey conducted by Inktomi Corparation in the year 2000, it was found that 87% of about one billion web pages available on the Internet are in English (Kiiski, 2001).

The founder of Microsoft, Bill Gates and the founder of Web, Tim Berners Lee, are both English speaking people. Microsoft software, such as Microsoft Internet Explorer and Microsoft Netscape Navigator are the core of Internet surfing. Thus, the acronymn for world wide web which is WWW, is claimed by the English speakers as three English words ruling the world (Leiner, 2001).

Table 1.2 below shows the number of Internet users for three years from 1997-1999 for countries where English is the native language, where English is a second language and where English is a foreign language. From the statistics shown, it could be seen that within three years, the number of Internet users in English speaking countries have increased by an average of 150%, by an average of 175% in countries where English is the second language and by an average of 116% in countries where English is a foreign language.

Table 1.2.

Three Year Comparison of Internet Users from 1997-1999 among the English Speaking Countries and Non-English Speaking Countries

Internet usership in English speaking countries	**1997 (million)**	**1998 (million)**	**1999 (million)**
USA	40m	60m	98m
UK	4.3m	8m	15.2m
Singapore	0.5m	0.75m	1.2m

'Table 1.2 continued' **Internet usership in countries using English as a second language**	**1997 (million)**	**1998 (million)**	**1999 (million)**
Japan	11.5m	16.7m	20.4m
France	1m	3.5m	7.2m
India	0.3m	0.5m	2.5m
'Table 1.2, continued,			
Malaysia	0.6m	0.8m	1.2m
Singapore	0.5m	0.75m	1.2m
Indonesia	0.25m	0.3m	1.3m
Internet usership in countries where English is a foreign language			
China	0.4m	2.1m	7m
Russia	0.15m	0.18m	0.23m

Source : http://www.euromktg.com/globstats

There seems to be greater increase in Internet users in countries where English is the second or foreign language. However, it does not reveal the language used by these users. Four years ago, English was the main language of 80% of the Web, although it is the mother tongue of less than half the world population. Now, multilingual sites are on the increase. Yahoo has sites in a dozen tongues (Houston, 1996). As stated by Salverda (2002), although English is useful as a lingua franca, it might not be enough at times for direct personal contact.

When web pages in multiple languages for two years are compared as shown in Table 1.3, it can be seen that other language websites have increased and English websites are on the decline. In fact the Guardian (ITAR, 2001), predicted that by the year 2005, Chinese will be the dominant language on the Internet as the webpages in Chinese are increasing.

Table 1.3 in the next page, shows that although English still dominates the Internet, other language websites are increasing at a fast pace. This could be due to technological advancement especially in the arena of software and operating systems. There have been efforts to provide software to support multilingual scripts. Such efforts have been initiated way back in 1998 by the Asia-Pacific Network Group (APNG). At present, multilingual technologies supporting DNS (Domain Name Systems) responsible for IP addresses (such as http://www....) have emerged, some of which were implemented by Yahoo.com. Among the first languages with multilingual scripts support in the DNS are Chinese, Japanese, Korean and among the Indian languages, Tamil (Venkata, 2001). The emergence of other languages on the Internet is not just a question of technology, but rather the interest and effort expended by the various language groups to promote their language on the Internet.

Table 1.3

2 Year Comparison of Pages in Multiple Languages .

	1996	2001
English	80%	49.9%
Chinese	1%	7.6%
Japanese	4%	7.2%
German	1%	5.9%
Spanish	1%	5%

Source : http://learnthenet.com/English

1.2.2. Tamil language on the Internet

Among the Indian languages, Tamil seems to have a significant number of websites on the Internet. A search using keywords 'Tamil language' and 'websites' provided links to 13,800 websites. However, on further probing, it was realised that not all these links and pages were in Tamil . Certain news sites had the feature 'translate' and when it is clicked, it displays the Tamil script. This is possible only if the Tamil font has been downloaded successfully in the computer and the font to read has been changed to read the Tamil script. Most of the links had homepages in the English language (Venkata, 2001).

There have been many efforts worldwide to develop websites in Tamil. An electronic English-Tamil dictionary has been developed by the South Asia Regional Studies, University of Pennsylvania (Wissink, 2001). In Malaysia, efforts have been taken to make available two literary works;

'Kurunthokai' and 'Thirumanthiram' on the web. The text files were converted to HTML (hypertext markup language) files and put up on the World Wide Web, so that it may be downloaded by anyone interested. Murasu Internet fonts which use the Romanised keyboard was used to type the literature. Microsoft Word intrinsically does not support Murasu fonts . It has to be downloaded from the Murasu site and loaded in Microsoft word (Wissink, 2001) Unicode implemented on softwares have been adopted for the Indic-languages. Unicode provides a unique number for every character, no matter what the program or the language is. Data, both in English and Tamil, will be stored in a UNICODE compliant database and will be retrieved in XML data. Different techniques will be used to display for different clients. HTML is all about presentation and is confined to web-browsers only, where as XML separates the data and the presentation. Since XML is fully UNICODE compliant, it enables easy and safe handling of non-Latin characters including Indian languages (Wissink, 2001).

With software and operating systems available now on the Internet to support Tamil fonts, producing technology texts in Tamil has become relatively easy.

1.3. Statement of the problem

While the information superhighway represents an evolutionary leap in communications hardware and software, the capacities of individual

users have not similarly evolved . Who are the real users of the Internet? Early internet user studies reveal that white males from higher socio-economic backgrounds are more likely to use the Internet (Kiiski, 2000). In June 1998, the New York Times reported that the average foreign Internet user will shift from highly educated, more affluent to moderate-income users who favor their native language (New York, 1998).

Tables 1.1 and 1.3 show that although other languages are on the increase on the Internet, English still remains the main language. One of the main reasons for this is the ease of technology which has a bias in favour of English. As all computers accept the ASCII code, English still remains the easiest language to surf the Internet. Guedon (1997), however predicts that despite its great world –wide strength, English will probably never emerge as the sole and universal lingua-franca. Guedon correctly foresees new possibilities of multilingual publishing in the next century with the developments in science and tecnology. Chinese, Spanish and Arabic users will improve their technology and accelerate their role to ensure that their languages are not left out on the Internet.

Given the ease of use of the English Language on the Internet where English is an inherent feature and where technological advancement now also supports non-Roman scripts, how do non-native speakers of English communicate with each other on the Internet? To address this statement, a website organised by and for non-native speakers of English

has been selected to analyse the language used to provide information and to communicate on the Internet.

1.4. Objective of the study

Given the importance of English language on the Internet at the moment and the increasing role of the other languages, the purpose of this study is to find out how much English is used in this chosen website on Tamil language where the participants are non-native speakers of English.

The objective of this study is also to find out the types of English and Tamil words used in this website. The emails archived in the website would also be analysed to find out if there is a pattern of code-switching from English to Tamil and vice versa. In relation to this, this study seeks to answer the following research questions :

Research Question 1 : How much English was used in this website in a month?

Research Question 2 : What are the types of English and Tamil words used in this website?

Research Question 3 : Is there a pattern of code-switching in the emails of this website?

1.5. Significance of the study

Many studies have been carried out on the kind of English used and the language choice among participants in the real world. In the virtual world like the Internet, there has been research done to study the English language used in the protocols of communication on the Internet such as emails, message boards and Internet Relay Chat. However few studies have been undertaken to study the choice of language in bilingual or multilingual websites.

Since technology now supports many languages and Internet users have a language choice when using the Internet, this study attempts to look into the language choice between English and Tamil in this chosen website on Tamil Language. This might be of use to future researchers interested in doing similar research. It might also provide insights into the sociolinguistics of multilingualism and bilingualism on the Internet.

1.6. Limitations of the study

Only one website for Tamil Language was chosen for this study. The other websites on Tamil Language were not considered. Though this sample website may not be typical of the population representing websites for Tamil Language, it has provided a starting point to examine language use in such websites.

Furthermore, not all portions of the sample chosen have been studied. Only the homepage and the emails from the archive for a certain time period have been analysed. Other sources of data in this website, such as projects and frequently asked questions, will not be considered. This is because using all the data in this website would amount to a huge data collection which may not be feasible to analyse for the purpose of this study.

The jetlag speed at which the cyberworld is progressing may make this study quite redundant. On-going studies should be continously carried out to analyse the amount of English used in websites on foreign languages.

1.7. Conclusion

This chapter provided a brief introduction to the languages used on the Internet and how English still dominates the Internet despite a rise in the use of other languages on the Internet. The following chapter will discuss similar studies carried out which have helped in the analysis of the research questions of this study.