# DISCUSSION PAPER SERIES

No. 3633

## THE EFFECT OF NATIVE LANGUAGE ON INTERNET USAGE

Neil Gandal

***INDUSTRIAL ORGANIZATION***

Centre for Economic Policy Research

www.cepr.org

**Available online at:** **www.cepr.org/pubs/dps/DP3633.asp**

*ISSN 0265-8003*

# THE EFFECT OF NATIVE LANGUAGE ON INTERNET USAGE

**Neil Gandal,** Tel Aviv University and CEPR

Discussion Paper No. 3633
November 2002

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 3633

November 2002

# ABSTRACT

## The Effect of Native Language on Internet Usage*

In this Paper, I explore the relationship between native language and use of the Internet and examine whether English is likely to retain its first-mover advantage of a large installed base of English language websites.

I study this issue empirically using a unique dataset on (home) Internet use at the individual level in Quebec from Media Metrix. The results suggest that English language websites are less of a barrier for French-speaking youths than for French-speaking adults in Quebec. To the extent that the younger generation drives the dynamics of the Internet, the results provide some support for the hypothesis that English will retain its first mover advantage of a large installed base of English language websites.

I also examine the effect of bilingualism on Internet use and find that among native French speakers, bilingual individuals use English language websites significantly more than their monolingual counterparts.

JEL Classification: D12 and L86
Keywords: internet use, language and network effects

Neil Gandal
Department of Public Policy
Tel Aviv University
Tel Aviv 69978
ISRAEL
Tel: (972 3) 640 6742
Fax: (972 3) 640 7382
Email: gandal@post.tau.ac.il

*I am grateful to the Neaman Institute and Haas School of Business at the University of California-Berkeley for financial support. I thank Carl Shapiro and seminar participants at the 3rd CEPR Conference on Applied Industrial Organization, the University of Arizona, the Federal Communications Commission, the Santa Anna School of Advanced Studies, and Tel Aviv University for helpful comments. I am grateful to Hanoch Frankovits, Jonathon Mi, Ajay Sreekanth, and Raymond Tchang for helpful research assistance, to Media Metrix for providing the data, and to Gregory Grefenstette and Clairvoyance.com for language identification analysis.

Submitted 29 September 2002

## 1. Introduction

In recent years, English has become the *de facto* standard for business and academic communication and has to some degree attained the status of a global language. English is the official language of the Asian trade group ASEAN and the official language of the European Central Bank, despite the fact that the bank is in (Frankfort) Germany and neither the U.K. nor Ireland are members of the European Monetary Union.[1] Several public schools in Zurich, Switzerland are now teaching some elementary school subjects in English. This is occurring in a country where there are four official languages --French, German, Italian, and Romansch. In a recent European Union survey, 70 percent agreed with the notion that everybody should speak English.[2]

In this paper, I examine how native language affects Internet use. The goal is to determine whether the Internet is likely to remain disproportionately English. Currently there is much more Internet content available in English than in other languages. A March 2001 estimate by Global Reach indicates that approximately 70 percent of all Internet content is in English. Japanese and German follow with approximately 6 percent each.[3] The Internet certainly is an effective instrument for circulating English around the world.

On the other hand, it is quite possible that English's first mover advantage (in terms of Internet content) will diminish: Although 40 percent of current Web users were native English speakers in March 2002,[4] web use is currently growing faster among non-native English speakers. Additionally because of low transaction costs, the Internet is ideal for bringing together members

---

[1] Wallraff, B., What Global Language, *Atlantic Monthly*, November 2000, p.53-66.

[2] Daley, S., "In Europe, Some Fear National Languages are Endangered," NY Times, 4/16/01.

[3] These estimates are virtually unchanged from a 1999 survey by "ExciteHome" that estimated that approximately 72 percent of all Internet content was in English, 7 percent was in Japanese, and 5 percent was in German. It should be noted, however, that these numbers are not precise, since they are based on surveys. Another 2001 estimate by the NEC Research Institute estimated that 86 percent of all Internet content was in English.

[4] Source: See Global Reach at http://www.glreach.com/globstats/index.php3. Following English, 10 percent of all Internet users are native Chinese speakers, while 9 percent are Japanese, 7 percent are Spanish, and 7 percent are German.

of small groups like speakers of Frisian, which is spoken by approximately 500,000 people throughout the world.[5]

Concern that national languages are becoming endangered might lead policy makers to require websites to be in the domestic national language. France already has many laws in place that protect the French language. Quebec requires all websites in that province be available in French. In Brazil, which has the largest Internet industry in South America, a bill was recently introduced that would prohibit the introduction and use of foreign words.[6]

### 1.1 Network Effects and Language

A key determinant of whether English will retain its first-mover advantage (of significant web content in English) is whether non-English speakers will use English language websites. In such a case, existing content providers would have little incentive to offer non-English versions of their sites, and new sites would have a strong incentive to provide their content in English. Such a first mover advantage may lead to a bandwagon because there are network effects in language: learning a second language is more valuable, the more widely that language is used.

A network effect exists when the value that consumers place on a particular product increases as the total number of consumers who use identical or compatible goods increases. In the case of an actual (or physical) network, such as the telephone or email network, the value of the network depends on the total number of subscribers who have access to the network. Since languages are in part communication technologies, the value of a language network increases in the number of speakers and users of that language. Languages are perfect substitutes, but they are incompatible in the sense that two individuals can talk with each other only if they both speak the same

---

[5] See Wallraff, B., "What Global Language," *Atlantic Monthly*, November 2000, p.53-66, and Nunberg, G., "Will the Internet always speak English?" *American Prospect*, March 27-April 10 2000 (available at www.prospect.org/print/V11/10/nunberg-g.html) for stimulating discussions of the issues.

[6] NY Times, 5/14/01, "English is Spoken Here Too Much, Some Say," by Larry Rohter.

language.[7] Languages, as a type of communications network, clearly are subject to strong direct network effects.

In the case of virtual networks, that are not linked physically, a network effect arises from positive feedback from complementary goods. The positive feedback mechanism works as follows: the value of the base product (such as a DVD player) is enhanced as the variety of (compatible) complementary products (content available on DVD discs) increases; hence consumers will be more likely to purchase a base product with many compatible complementary products. The variety of complementary products, in turn, will depend on the total number of consumers that purchase the base product. As the number of consumers that purchase the base product increases, there is a greater demand for compatible complementary products. This increases the profitability of supplying complementary products. Since there are typically fixed or sunk entry costs, production of the complementary products is characterized by increasing returns to scale. So, more complementary products will be produced or developed for a base product with a large share of the market. This further enhances the value of the base product, causing positive feedback in the system: an increase in the sales of the base product leads to more compatible complementary products, which further increases (the value of and) sales of the base product. See Chou and Shy (1990) and Church and Gandal (1992).

Languages are also subject to virtual network effects. The value of speaking English increases as the number of English language websites (or other content, such as books, magazines, or movies) increases. This will lead to an increase in the number of non-English speakers learning English in order to have access to English language websites, since individuals who speak English will have more websites to use. This in turn will lead to an increase in the number of English language websites.

Markets in which there are network effects are often characterized by tipping: once a system has gained an initial lead, there is a snowball effect. Katz and Shapiro (1994, p. 105) note that positive feedback means that there is a "natural tendency towards de facto standardization."

---

[7] Of course, when there is no common language, people can still communicate in non verbal ways, such as gestures, expressions, etc.

Hence it is possible that the first-mover advantage of English may result in English remaining the dominant language on the web, even though there are more native Chinese and Spanish speakers than there are native English speakers. The outcome will depend, in large part, on the strength of the virtual network effect: Does the large number of English language websites encourage non English speakers to learn English so that they can access them?[8]

The use of language on the Internet can fruitfully be viewed as a *co-adoption process*.[9] Here "adoption" means use of a particular language; thus *language training and use* are comparable to *technology adoption* decisions that have been extensively studied. The operator of a website "adopts" a language by offering its site in that language. Likewise, an individual "adopts" a language by learning that language.

More specifically, focusing on the *decisions* made by websites and users, one can examine the dynamics of language adoption over three time frames. In the *short-term* (day to day), individuals decide – based in part on their language skills and in part on the available offerings in different languages – which websites to visit, how long to stay at these sites, and whether to engage in commercial transactions. These decisions determine actual Internet usage by different individuals and groups (such as the group of native French speakers). In the *medium term*, operators of websites decide which language to use for their site, and whether to offer their sites in multiple languages (if permitted this choice by their local governments). These decisions are driven in large part by the amount of traffic that a site expects to attract in one language or another, plus the *incremental traffic* that a site expects to attract by offering its content in multiple languages. Over the *long term*, individuals (and their parents and teachers) make decisions about which languages to learn. This decisions are driven in part by the desire to

---

[8] In the case of language, translation is an ex-post substitute for compatibility. If translation utilities worked well, the issue of language would likely be less important. Given the subtleties involved in language, translation by artificial intelligence is in its infancy and works quite poorly.

[9] Co-adoption processes are common when virtual network effects are present. For example, when there were two rival incompatible formats for 56k modems, Internet Service Providers and consumers selected modem formats, each influenced by the other group's decisions. This same dynamic arises as well in various server client architectures, and influenced the battle between the Netscape Navigator and Microsoft Internet Explorer browsers, as well as the adoption and use of Sun's Java.

access certain content, as well as the desire to communicate directly with others speaking other languages.

### 1.2 A Road Map

This paper focuses on the short-term behavior. In order to examine this issue empirically, I have obtained a unique data set on (home) Internet use at the individual level in Quebec, Canada from Media Metrix. Canada provides an ideal setting to examine this issue because English is one of the two official languages. French is, of course, the other official language of Canada. Quebec is ideal because of the fact that there are significant numbers of both native French and native English speakers.[10] The attractiveness of beginning with a single country (and a single region within a country) is that there is typically greater heterogeneity across countries (or regions) than within a single country (or a single region.)[11,12]

I describe the data in section 2. Empirical results are provided in Section 3. The results suggest that English language websites are less of a barrier to Internet use for French-speaking youth than for French speaking adults in Quebec. To the extent that the younger generation drives the dynamics of the Internet, the results provide some support for the hypothesis that English will maintain its first mover advantage (of a large installed base of English language websites) on the Internet. Among native French speakers, bilingual individuals use English language websites significantly more than their monolingual counterparts. This result is quite similar across different age groups. Section 4 briefly concludes. I close this section with a brief overview of the literature.

### 1.3 Brief Literature Review

---

[10] Indeed, many of the studies in the economics of language focus on Canada in general, and Quebec in particular. See Grin and Vaillancourt (1997).

[11] For example, in some countries, local phone calls are metered, while in other countries, there is a fixed monthly charge for local service. Additionally, Internet access speed might differ widely by country. It can be difficult to find data on and control for these variables.

[12] But the techniques discussed are applicable to other regions of the world. There is some casual evidence that the most popular online destinations in Mexico originate in the U.S. and do not target Spanish Speakers. See Heft, D. "Who Rules the Internet in Mexico. Why It's America," The Standard, June 12, 2001, at http://www.thestandard.com/article/0,1902,27096,00.html.

This work complements the growing literature on the economics of language. Grin and Vaillancourt (1997) provide an overview of the literature; a nice survey is provided by Grin (1996). The major research area within this field is the empirical relationship between earnings and language attributes. Several recent papers are Bloom and Grenier (1996), Chiswick and Miller (1999), and Zavodny (2000). Grin (1990) and Church and King (1993) examine rational language choice and public policy toward bilingualism using theoretical models. Rauch (1999) shows that common language facilitates international trade in differentiated products. Bertrand, Luttmer, and Mullainathan (2000) find that for high welfare-using language groups, being surrounded by speakers of the same language leads to higher welfare use. Freund and Weinhold (2000) find that increased access to the Internet increases trade flows among developed countries. To the best of my knowledge, there is no work on the relationship between Internet use and language, which is the focus of this study.

# 2. Data

## 2.1 Description of Data

I employ a unique data set on Internet use at the individual level in Canada, which comes from Media Metrix, the industry leader in the measurement of Internet use. The data include information on demographics of the user such as income, education, family size, province, etc. The data on Internet use is very detailed. Complete click-stream data are available for the December 2000 period. These data include a separate entry for each URL that is visited, as well as the "active time" spent at each URL location.[13] Additionally, and this is key for the study, the mother tongue of the user – English or French – is known.

The study is restricted to Quebec. The reason for doing so is that there may be significant differences among provinces on variables for which data are not available, such as the speed of Internet service. Hence, it makes sense to look at Quebec, which is the only province in Canada

[13] If a user does not make an entry for 60 seconds, the active time count is halted. Hence for less than sixty seconds at a website, active time is equal to total time. For time spent on a page beyond 60 seconds, active time is less than or equal to total time. Data on total time are available as well.

with significant proportions of both native English and native French speakers in the Media Metrix sample.[14]

The sample is at the level of the individual user and for reasons that will become immediately apparent, the sample is restricted to households with multiple users. The data include observations on 490 individuals from the age of 10 through 70 inclusive from 199 households;[15] 410 users are native French speakers, while 80 users are native English speakers.

In order to examine cohort/generation effects, I break the sample into three groups: ages 10-20, ages 21-45, and ages 46-70. Of the 80 English speakers in the sample, 26 were in the 10-20 age group, 33 were in the 21-45 age group and 21 were in the 46-70 age group. Of the 410 native French speakers, 101 were in the 10-20 age group, 211 were in the 21-45 age group and 98 were in the 46-70 age group.

The following variables are available for the study:

- Active Time – This is the total time (in hours) that the user was active on the Internet during the December 2000 period.[16]
- Active Time of Others – This is the total time (in hours) that other members of the household were active on the Internet during the December 2000 period.
- Age – Age of the user
- Female – A dummy variable that takes on the value 1 if the user is female and takes on the value 0 if the user is male.

---

[14] According to Statistics Canada (1996), there are 602,865 native English speakers in Quebec and 5,728,290 native French speakers in the province. There are 50,585 people who are both native French and English speakers.

[15] There are younger users in the sample, but they are more likely looking at images, rather than using language skills.

[16] Since there is a separate entry for each page visited, one could use the total number of pages visited rather than active time. I believe, however, that active time is a better measure of the importance of each "visit. Fully 1/3 of the observations account for more than 80 percent of active time.

- Language – A dummy variable that takes on the value 1 if French is the mother tongue of the user and 0 if English is the mother tongue of the user.

- Size – Equal to the number of members of the household, up to a maximum of five. All households with 5 or more members have size equal to five.

- Income – The variable takes on the value 1 if the household income is less than $25,000, 2 if household income is between $25,000 and $40,000, 3 if household income is between $40,000 and $60,000, 4 if household income is between $60,000 and $75,000, 5 if household income is between $75,000 and $100,000, and 6 if household income exceeds $100,000.

- Kids – This is a dummy variable equal to 1 if there are children under age 18 in the household; the dummy variable is equal to 0 if there are no children under age 18 in the household.

- HS_grad – The (dummy) variable takes on the value 1 if the individual has completed high school, but does not have a college degree. Otherwise, the variable takes on the value zero. (This variable is only employed for individuals who are at least 21 years old.)

- College_grad – The variable takes on the value 1 if the individual has a college (B.A.) degree or a higher degree. Otherwise, the variable takes on the value zero. (This variable is only employed for individuals who are at least 21 years old.)

- Pereng – This variable is defined to be the percent of the active time that was spent on websites whose content is in English out of the total active time spent at websites whose content is either English or French.[17]

- Fbilingual – This variable takes on the value 1 if native French speakers come from households where other household members spend on average 80% or more of their

[17]Data were not collected on the language of the website. See the discussion below.

active time at English language websites. Otherwise, this variable takes on the value zero.[18]

### 2.2 Creation of the Variable *Pereng*

Data were not collected on the language of the website. Although, there were more than 1.3 million URL "full pages" in the sample, there are "only" approximately 40,000 unique URL domains. (An example of a URL domain is http://www.sfgate.com and an example of a URL full page is http://www.sfgate.com/classifieds/rentals/.) Given the prohibitive expense required to purchase language identification services to process 1.3 million URLs, a simple computer (spider) program was written to determine the language of each of the approximately 40,000 unique URL domains in the sample.[19] Hence, by this method, I am able to assign a language to websites. This is, of course, an approximation since some of the URL domains contain some "interior" pages that are in one language and some interior pages that are in other languages. Nevertheless the assumption is probably quite reasonable for my purposes.

The spider program classified all unique URL domains that have ASCII characters above 192 as French. This includes all characters with accents marks, such as "é" and "û". If such characters were not present, the website was characterized as English. This is also an approximation. In order to examine the approximation, a separate language identification program examined approximately 100,000 URL full pages in the sample. This analysis indicated that less than 3% of the URL full pages in the data set were in languages other than English or French.[20] Hence, this approximation also seems quite reasonable.

Fully 80 percent of all active time was characterized by the spider program. The 20% of active time for which the language was not identified is due to the following reasons: (i) the unclassified websites were services that required the user to enter his/her personal ID or

---

[18] I chose 80 percent, because it is close to but slightly less than the average percent of time that a native English speaker spends on English language websites. According to this definition, 29% of native French speakers in the sample from households with multiple users are "bilingual." According to Statistics Canada, 32% of all native French speakers in Quebec are bilingual.

[19] The spider program is not able to handle massive numbers of URLs.

information, (ii) the website redirected the user more than four times, or (iii) the website did not exist when the data were analyzed.

### 2.3 Web Site Categories

It is important to categorize the "type" of website accessed, so we can understand in greater detail how different types of Internet usage are influenced by language. Using their category definitions, Media Metrix classified approximately 25 percent of the 40,000 URL domains. Using the Media Metrix definitions, research assistants completed the task for the other 30,000 URL domains. The categories are as follows:[21]

1. Retail, Business
2. Information (News, Entertainment, Sports)
3. Education
4. Search (Portals, Directories, Reference)
5. Services (ISPs, Discussion/Chat, email)
6. Government
7. Adult

### 2.4 Descriptive Statistics

Descriptive and summary statistics are contained in Table 1. The table shows that, overall, native English speakers spent approximately 26 percent more time on the Internet than native French speakers. Table 1 also shows that, on average, native English speakers in Quebec accessed English content websites 88 percent of the time, while native French speaking Quebecois accessed English content websites 65 percent of the time. This already suggests that Quebecois are using the web intensively in English.

Figure 1 breaks down active time on the Internet by age group and native language. The figure shows that there are differences among age groups as well as among English and French speakers. In the youngest age group of users in the analysis (ages 10-20), native French speakers spent approximately 19 percent more time on the Internet than native English speakers, while in

---

[20] I am extremely grateful to Gregory Grefenstette and Clairvoyance.com for conducting this analysis and providing these results. See Grefenstette (1995) for an overview of language identification software programs.

[21] Media Metrix had 26 categories. I've aggregated their categories into seven broader categories.

the 21-45 age group, native English speakers spent approximately 74 percent more time on the Internet than native French speakers. There is no difference in usage for the 46-70 age group.

Figure 2 delineates active time by category and native language. Overall, users spend 32 percent of their active time at search/portal sites, followed by retail/business (22 percent of active time) and services (21 percent of active time). Users spend approximately three percent of active time at government and educational sites combined.

Figure 2 shows that native French speakers spend 2.2 percent of their active time using websites in the government category, while native English speakers spend only 1.1 percent of their time using websites in the government category. Since all government websites in Canada are in both English and French, this suggests that native French speakers have a preference for websites in their native language.[22] Figure 2 shows that native French speakers spend a greater percent of their active time using websites in the services category than native English speakers; these sites (email services, chat/discussion) are also likely to provide information in both English and French.

Figure 3 breaks down active time by age, category and native language. The figure shows that the younger users (ages 10-20) spend much less time at Business/Retail websites than users in the other two age groups. Figure 3 also shows that native French speakers in the 21-45 age group spend 2.6 percent of their active time using websites in the government category; native English speakers in the same age group spend only 1.0 percent of their time using websites in the government category. This suggests that native French speakers in the 21-45 age group indeed have a preference for websites in their native language.

## 3. Empirical Results

The empirical work focuses on the critical short-term behavior. As mentioned above, short-term behavior is crucial in determining whether English will maintain its first-mover advantage on the Web. I now describe the empirical tests that are conducted.

### 3.1 Explaining the Percent of Time Spent on English Language Websites

I first examine whether there are differences between native French and English speakers regarding the percent of the time that each user spends at English language websites.

I run the regressions by age group because different age groups typically use the web for different purposes and the availability of websites in French and English may differ by category. (Unfortunately, data on availability of websites by category and language are not available.) Additionally, I am interested in whether there are generational/cohort effects that may be associated with the adoption/use of the Internet.

According to Statistics Canada (1996), 32 percent of all native French speakers in Quebec claim knowledge of English.[23] Hence, Quebec has a population with many bilingual individuals. Since the data do not contain information on bilingualism, I attempt to control for bilingualism of native French speakers, by including the variable Fbilingual as an explanatory variable. As long as bilingual ability is relatively equal among household members, this variable will reasonably control for the degree of bilingual ability of the user.

It is not appropriate to run ordinary least squares (OLS) regressions with the percent of the time that each user spends at English language websites (*pereng*) as the dependent variable. This is because *pereng* only takes on values between 0 and 1. Given that *pereng* ranges between 0 and 1, "logit" regressions are employed in table 2. For ease of presentation, let $p_i$ be the percent of time that an individual spends at English language websites. The regression equation can be written as

$$\log[p_i/(1-p_i)] = \beta_0 + \beta_1 x_{1\,i} + \beta_2 x_{2\,i} + \ldots + \beta_k x_{k\,i} + \varepsilon_i ,$$

where i refers to the individual, $\beta_0, \beta_1, \ldots, \beta_k$ are the coefficients to be estimated and $x_1, x_2, .., x_k$ are the independent variables. We've added "one second" of time to total time in English and

---

[22] Another explanation is that native French speakers are more politically active and hence spend more time at government websites.

[23] The official definition is "knowledge of official languages." Fully 40 percent of all native English speakers in Quebec claim knowledge of French as well.

total time in French, so that observations are not lost. This doesn't distort use, because on average, individuals spend more than eight hours a month on the Internet. I do this because 10 individuals in the sample access the web only in French, while 31 individuals in the sample access the web only in English. If we did not add this "one second of time" the dependent variable, $\log[p_i/(1-p_i)]$, would be undefined for these individuals.[24]

I run a weighted least squares (WLS) regression, where the weights come from the assumption that each second is an independent observation from a Bernoulli distribution. Since each individual accesses a very large number of websites, this is a reasonable approximation.

Table 2 shows that native French speakers spend significantly less time at English language websites than native English speakers. Although this is true for all age groups, the effect is much stronger for the 21 to 45 age group and the 46 to 70 age group. That is, English language websites are less of a barrier for the youngest group (age 10 to 20) of native French speakers.

The table also shows that native French speakers who are bilingual use English language websites more than native French speakers who are not bilingual. The effect is strikingly similar across all three age groups. For the youngest age group, there is little difference between native English speakers and bilingual French speakers in terms of use of English language websites.[25]

### 3.2 Explaining Active Time Spent on the Internet

In Table 3, OLS regression results with the logarithm of active time in hours as the dependent variable are presented.[26] Regressions are run by age group, using the same categories as above. I include all of the demographic variables that were included in Table 2 as explanatory variables

---

[24] In any case, the estimates are very robust to dropping the individuals when $p_i$=1 or 0.

[25] I ran the same regressions reported in table 2 again under the assumption that only observations from different households are independent. This is done by using the "robust" option in Stata. The parameter estimates, are, of course, unchanged. The estimated t-statistics are somewhat smaller, but language and fbilingual remain statistically significant. Similarly, I re-estimated the regression in table 3 under the assumption that only observations from different households are independent. There is little change in this case. These results are available from the authors upon request.

[26] I employ the logarithm because of the large dispersion of active time. But the results are qualitatively unchanged if active time is used as the dependent variable.

as well as the logarithm of the active time of others in the household as a regressor in order to examine whether use by other household members affects an individual's Internet use.

The results in table 3 suggest that native language is not a barrier to Internet use for either the youngest group of native French speakers (ages 10-20) or the oldest group of native French speakers (ages 46-70). For the 21-45 age group, the estimated coefficient on language is negative and statistically significant, suggesting that, in this age group, native French speakers spend less time on the Internet than native English speakers. Table 3 shows that in general, the other explanatory variables do not explain much of the difference in active time across individuals.

## 4. Conclusion

The results in section 3 suggest that English language websites are not that significant a barrier for French-speaking youth in Quebec. To the extent that the younger generation drives the dynamics of the Internet, the results provide some support for the hypothesis that English will retain its first mover advantage on the Internet. A key question is whether this is a cohort effect, i.e., an effect that will persist over time even as the cohort ages, or a generational effect. If it is the latter, then it is likely that English will not retain its first mover advantage, since as the youngest users age, their Internet use will reflect that of the 21-45 age group. Native French speakers in this age group spend significantly less time using the Internet and significantly less time at English language websites than their native English speaking counterparts. If the 21-45 age group drives the dynamics of web use, the analysis suggests that operators of websites will likely find it worthwhile to offer their sites in multiple languages; in such a case, English will not retain the first-mover advantage of a large installed base of English language websites.

# References

Bertrand, M., Luttmer, E., and S. Mullainathan, 2000, "Network Effects and Welfare Cultures, " *Quarterly Journal of Economics*, 115: 1019-1055.

Bloom, D., and G. Grenier, "Language, Employment, and Earnings in the United States: Spanish-English Differentials from 1970 to 1990," *International Journal of the Sociology of Language*, 121, 45-68, 1996.

Chiswick, B., and R. Miller, "The 'Double Negative' Effect on Earnings of Limited Language Proficiency Among Immigrants in Canada," University of Chicago Center for the Study of the Economy and the State Working Paper: 148, February 1999.

Chou, C. and O. Shy, 1990, "Network Effects without Network Externalities," *International Journal of Industrial Organization*, 8: 259-270, 1990.

Church, J., and I. King, "Bilingualism and Network Externalities," *Canadian Journal of Economics*, 26(2):337-345, 1993.

Church and Gandal, "Network Effects, Software Provision and Standardization," *Journal of Industrial Economics*, XL: 85-104, 1992.

Freund, C., and D. Weinhold, "International Trade and the Internet: Some Preliminary Evidence," mimeo, 2000.

Grefenstette, G., 1995, "Comparing Two Language Identification Schemes," 3rd annual conference on Statistical Analysis of Textual Data.

Grin, F., "The Economic Approach to Minority Languages," *Journal of Multilingual and Multicultural Development*, 11, 153-171, 1990.

Grin, F., "The Economics of Language: Survey, Assessment and Prospects," *International Journal of the Sociology of Language*, 121, 17-44, 1996.

Grin, F. and F. Vaillancourt, 1997, "The Economics of Multilingualism: Overview of the Literature and Analytical Framework", in W. Grabe (ed.), Multilingualism and Multilingual Communities. Cambridge: Cambridge University Press, 43-65.

Katz, M., and C. Shapiro, "Systems Competition and Network Effects," *Journal of Economic Perspectives*, 8: 93-115, 1994.

Rauch, J., "Networks versus Markets in International Trade," *Journal of International Economics*, 48(1): 7-35, 2000.

Zavodny, M., "The Effects of Official English Laws on Limited-English-Proficient Workers," *Journal of Labor Economics,* 18(3): 427-52, 2000.

# Appendix: Tables & Figures

## Table 1. Descriptive Statistics

| French speakers, N=410 | | | | |
|---|---|---|---|---|
| VARIABLE | MEAN | STD. DEV | MINIMUM | MAXIMUM |
| Age | 34.14 | 14.82 | 10 | 70 |
| Female | 0.50 | .50 | 0 | 1 |
| Income | 3.41 | 1.44 | 1 | 6 |
| Size | 3.17 | 1.14 | 1 | 5 |
| HS_grad | 0.29 | 0.45 | 1 | 7 |
| College_grad | 0.43 | 0.49 | | |
| Kids | 0.56 | 0.50 | 0 | 1 |
| Active time (hours) | 7.85 | 11.64 | .01 | 94.44 |
| Pereng | 0.65 | 0.26 | 0 | 1 |
| Fbilingual | 0.29 | 0.46 | 0 | 1 |
| English Speakers N=80 | | | | |
| VARIABLE | MEAN | STD. DEV | MINIMUM | MAXIMUM |
| Age | 31.86 | 15.08 | 10 | 70 |
| Female | 0.6 | 0.49 | 0 | 1 |
| Income | 3.65 | 1.71 | 1 | 6 |
| Size | 3.49 | 1.40 | 1 | 5 |
| HS_grad | 0.33 | 0.47 | 1 | 7 |
| College_grad | 0.31 | 0.47 | 0 | 1 |
| Active Time (hours) | 9.86 | 15.93 | 0.08 | 114.47 |
| Kids | 0.60 | 0.49 | 0 | 1 |
| Pereng | 0.88 | 0.18 | 0 | 1 |

**Figure 1: Average Monthly Internet Use**

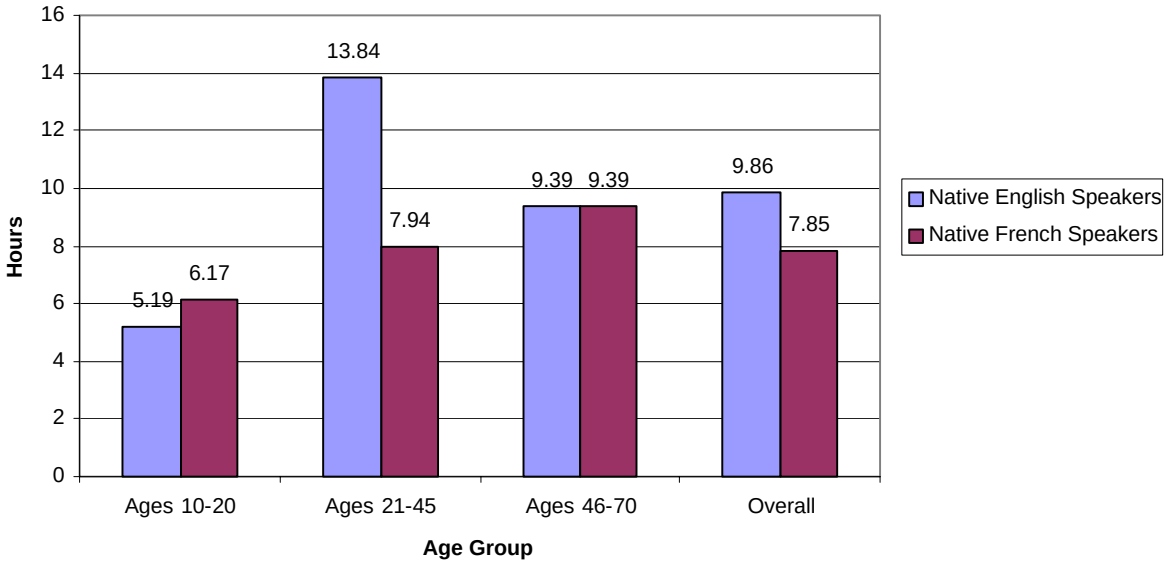
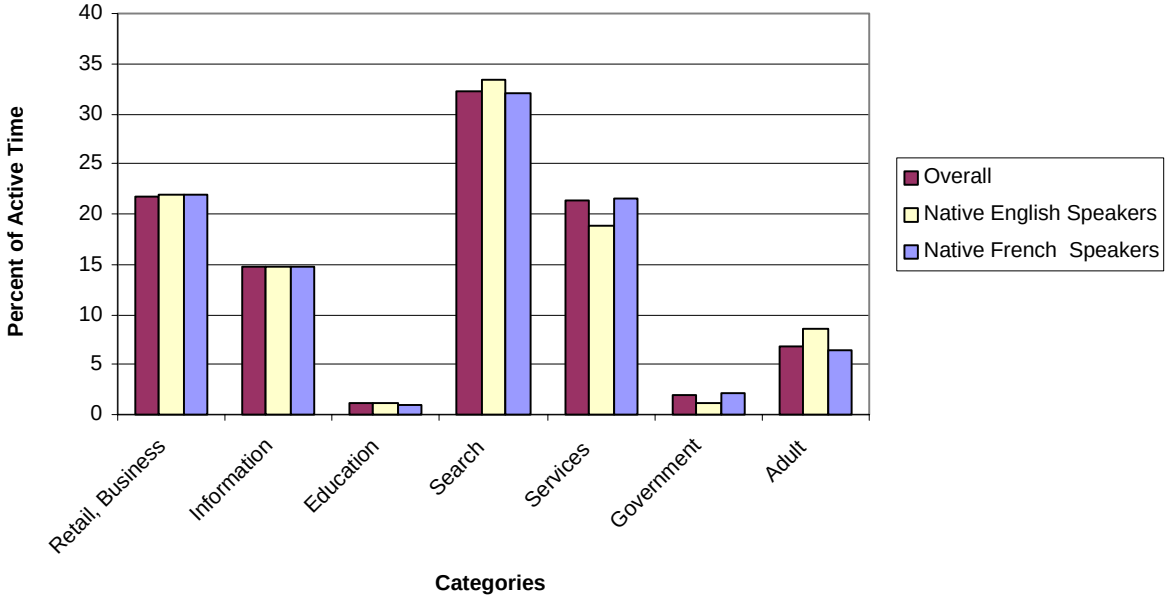


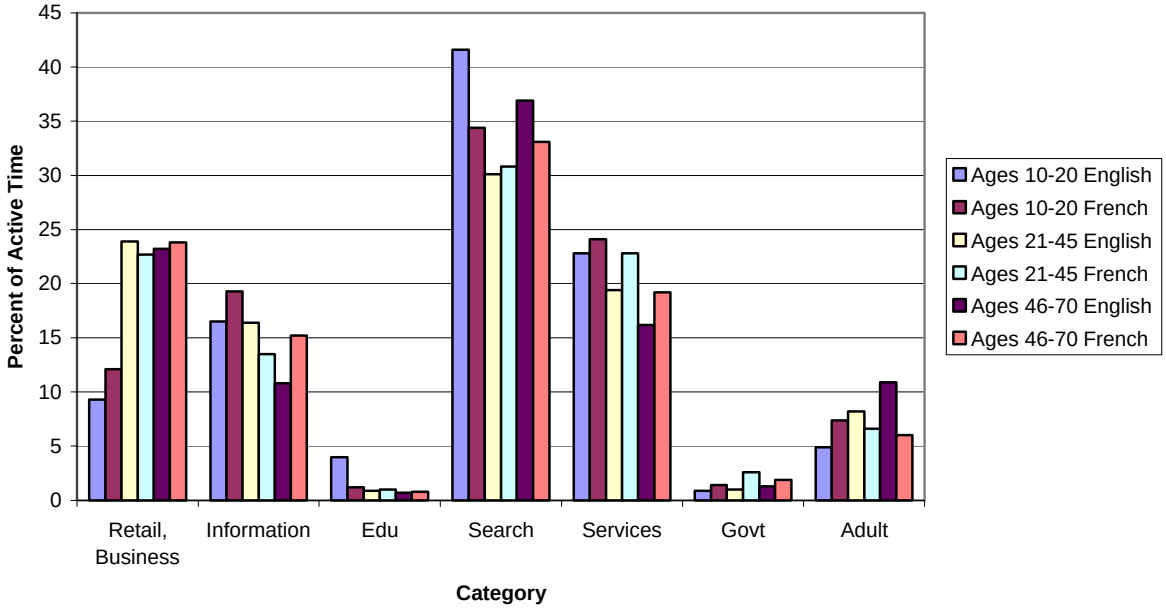
**Figure 2: Percent of Active Time by Category & Native Language**




# Percent of Active Time by Category & Language

| | Overall | English | French |
|---|---|---|---|
| Retail, Business | 21.8 | 22.0 | 22.0 |
| Information | 14.8 | 14.8 | 14.8 |
| Education | 1.1 | 1.2 | 1.0 |
| Search | 32.3 | 33.4 | 32.0 |
| Services | 21.3 | 18.9 | 21.6 |
| Government | 2.0 | 1.1 | 2.2 |
| Adult | 6.8 | 8.6 | 6.5 |

**Figure 3: Percent of Active Time by Age Group, Native Language, and Category**




# Percent of Active Time by Age, Language, and Category (E=English, F=French).

| | | Percent of Active Time By Category | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Age Group | | Retail, Business | Information | Education | Search | Services | Government | Adult |
| 10-20 | E | 9.3 | 16.5 | 4.0 | 41.6 | 22.8 | 0.9 | 4.9 |
| | F | 12.1 | 19.3 | 1.2 | 34.4 | 24.1 | 1.4 | 7.4 |
| 21-45 | E | 23.9 | 16.4 | 0.9 | 30.1 | 19.4 | 1.0 | 8.2 |
| | F | 22.7 | 13.5 | 1.0 | 30.8 | 22.8 | 2.6 | 6.6 |
| 46-70 | E | 23.2 | 10.8 | 0.7 | 36.9 | 16.2 | 1.3 | 10.9 |
| | F | 23.8 | 15.2 | 0.8 | 33.1 | 19.2 | 1.9 | 6.0 |

## Table 2: Weighted Least Squared Regression Results

Dependent Variable is $\log[p_i/(1-p_i)]$. These regressions are done at the level of the individual by age group. The t-values are in parentheses.

| | Ages 10 to 20 | Ages 21 to 45 | Ages 46 to 70 |
|---|---|---|---|
| Constant | 2.55 (3.28) | 3.51 (7.60) | 2.26 (2.17) |
| Fbilingual | 0.59 (3.20) | 0.65 (4.37) | 0.63 (3.06) |
| Age | -0.057 (-1. 63) | -0.025 (-2.62) | -0.0029 (-0.18) |
| Female | -0.71 (-4.27) | -0.19 (-1.62) | -0.36 (-1.88) |
| Language | -0. 89 (-2.99) | -1.65 (-8.59) | -1.33 (-4.07) |
| Size | 0. 098 (1.22) | -0.10 (-1.61) | -0.16 (-1.25) |
| Income | -0.037 (-0.61) | 0.063 (1.46) | -0.024 (-0.32) |
| Kids | 0.16 (0.67) | 0.24 (1. 52) | 0.63 (1.60) |
| HS_grad | | -0.75 (-2.97) | 0.33 (1.02) |
| College_grad | | -0.50 (-2.00) | 0.026 (0.08) |
| # of Obs. | 127 | 244 | 119 |
| $R^2$ | 0.25 | 0.32 | 0.21 |
| Adj $R^2$ | 0.21 | 0.30 | 0.15 |

## Table 3: OLS Regressions with log(Active Time) as Dependent Variable

These regressions are done at the level of the individual by age group. The t-values are in parentheses.

| | Ages 10 to 20 | Ages 21 to 45 | Ages 46 to 70 |
|---|---|---|---|
| Constant | -0.76 (-0.79) | 1.76 (1.86) | 3.11 (1.88) |
| Age | 0.062 (1.34) | 0.0042 (0.17) | -0.019 (-0.79) |
| Female | 0.28 (1.014) | -0.15 (-0.64) | -0.67 (-2.49) |
| Fbilingual | 0.41 (1.25) | -0.089 (-0.33) | 0.0098 (0.03) |
| Language | 0.0004 (0.00) | -0.58 (-1.68) | -0.14 (-0.37) |
| Size | 0.015 (0.09) | -0.13 (-0.90) | -0.22 (-1.32) |
| Income | -0.15 (-1.57) | -0.034 (-0.40) | -0.041 (-0.36) |
| Kids | 0.89 (2.05) | 0.46 (1.46) | 0.53 (1.07) |
| Log(actimeothers) | 0.10 (0.92) | 0.10 (1.27) | -0.058 (-0.64) |
| HS_grad | | 0.19 (0.34) | 0.69 (1.45) |
| College_grad | | -0.24 (0.43) | 0.61 (1.24) |
| # of Obs. | 127 | 244 | 119 |
| $R^2$ | 0.12 | 0.05 | 0.09 |
| Adj $R^2$ | 0.06 | 0.01 | 0.005 |