Native Language and Internet Usage

Neil Gandal (Tel Aviv University, Michigan State University, and CEPR)

December 2003

Abstract

In this paper, I explore the relationship between native language and use of the Internet. I study this issue empirically using a unique data set on Internet use at the individual level in Quebec. The results suggest that most Quebecois are using the web intensively in English. Furthermore, the difference between native French and native English speakers in the use of the Internet in English is relatively small for the younger generation. To the extent that the younger generation drives the dynamics of the Internet, the results provide support for the hypothesis that English will retain its first mover advantage of a large installed base of English language websites.

JEL Classification Numbers: L86, D12

Keywords: Language, Internet Use, Network Effects

I thank the Neaman Institute and Haas School of Business at the University of California-Berkeley for financial support. I'm especially grateful to Shane Greenstein and I also thank Carl Shapiro and seminar participants at the 2003 American Economics Association Annual Meeting, 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.

I. 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 in English. Currently there is much more Internet content available in English than in other languages. A March 2003 estimate by Global Reach indicates that approximately 68 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 35 percent of Web users were native English speakers in March 2003,[4] web use is growing faster among non-native English speakers. Additionally because of low transaction costs, the Internet is ideal for bringing together members of small groups like speakers of Frisian, which is spoken by approximately 500,000 people throughout the world.[5]

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.[6] The attractiveness of studying 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.)[7]

The results suggest that most native French speakers (Quebecois) are using the web intensively in English. Furthermore, the difference between French and English speakers in the use of the Internet in English is relatively small for the younger generation. To the extent that the younger generation drives the dynamics of the Internet, the results provide support for the hypothesis that English will maintain its first mover advantage (of a large installed base of English language websites) on the Internet.

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. This paper is most closely related to the literature on the relationship between native language and international trade. Rauch [1999] shows that proximity and common language/colonial ties facilitate international trade in differentiated products. Freund and Weinhold [2002] find that increased Internet use is correlated with the growth in international trade in services. They are careful to note that while their "results imply that there is a correlation between Internet adoption and international services trade, they say little about causality (p.239)." The results of my paper suggest that the link between international trade and Internet use may be through the increased use of a common language, namely English.

This work also 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.

The relationship between language and the Internet use is discussed in section II. I describe the data in section III. Empirical results are provided in Section IV. Section V briefly concludes.

II. The Use of Language on the Internet: A Co-adoption Process

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 language. 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." 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 operators of 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. 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 access certain content, as well as the desire to communicate directly with speakers of other languages.

The empirical work in section IV focuses on the critical short-term behavior. The short-term behavior is crucial in determining whether English will maintain its first-mover advantage on the Web. In the following section I describe the data used for the study.

III. Data

III.A Description of Data

I employ a unique data set on Internet use at the individual level in Quebec, 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, 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.[10] 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 with significant proportions of both native English and native French speakers in the Media Metrix sample.

The sample is at the level of the individual user and for reasons that will become apparent, the sample is restricted to households with multiple users.[11] The data include observations on 490 individuals from the age of 10 through 70 inclusive from 199 households;[12] 410 users are native French speakers, while 80 users are native English speakers.

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.[13]

- 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.

- 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. Data were not collected on the language of the websites. See section III.B for discussion on the creation of the variable "Pereng."

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

I chose 80 percent, because it is close to but less than the average percent of time (88 percent) that a native English speaker spends on English language websites. According to this definition, 29% of native French speakers in the sample are "bilingual." According to Statistics Canada [1996], 32 percent of all native French speakers in Quebec claim knowledge of English.[14] 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.

III.B Creation of the Variable *Pereng*

Data were not collected on the language of the websites. 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[15] 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.[16] 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 because the spider program may be classifying other languages (such as German for example) as French. In order to examine this effect, a 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.[17] 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 information, (ii) the website redirected the user more than four times, or (iii) the website did not exist when the spider program was employed.

IV. 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. Hence I 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.

In section IV.A I examine summary statistics and in section IV.B I employ some simple regressions. In both sections I examine the data by age group because:

- I am interested in whether there are generational/cohort effects that may be associated with the adoption/use of the Internet.

- 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. Analyzing the data by age group controls for "unobserved" supply effects.

Hence I employ the following age groups in the analysis: 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.

IV.A Summary Statistics

Descriptive and summary statistics are contained in Table I. Table I shows that, on average, native English speakers in Quebec accessed English content websites 88 percent of the time, while native French speakers (Quebecois) accessed English content websites 65 percent of the time. This already suggests that, in general, Quebecois are using the web intensively in English.

Figure I delineates percent of time spent on English language web sites by age group and native language. The figure shows that there are differences among age groups as well as among English and French speakers.

The youngest users (ages 10-20) spent the most time at English language web sites. This is the case for both young English speakers -- who spent 91 percent of their time at English language web sites -- and young French speakers – who spent 72 percent of their time at English language web sites. Native French speakers in the 21-45 (46-70) age group spent 61 percent (64 percent) of their time at English language web sites while their English speaking counterparts in the 21-45 (46-70) age group spent 84 percent (89 percent) of their time at English language web sites.

Figure II shows that, overall, native English speakers spent approximately 9.86 active hours on the Internet per month, while native French speakers spent 7.85 active hours on the Internet per month. Thus, on average, native English speakers spend approximately 26 percent more time on the Internet than native French speakers. Figure II 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 the 21-45 age group, native English speakers spent approximately 74 percent more time on the Internet than native French speakers. There was no difference in usage for the 46-70 age group.

IV.B Explaining the Percent of Time on English Language Websites Using Regressions

In this section, I employ regressions to explain the percent of time spent on English language websites. 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 II. 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.

I've added "one second" of time to total time in English and 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 31 native English speakers 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.

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.

The regression results in table II shows that native French speakers spend significantly less time at English language websites than native English speakers in their cohort. Although this is true for all age groups, the effect is much stronger for the 21 to 45 age group (t=-8.59) and the 46 to 70 age group (t=-4.07). That is, English language websites are less of a barrier for the youngest group (age 10 to 20) of native French speakers (t=-2.99). Indeed the difference between native French speakers and native English speakers in the youngest group is smaller than for the other two groups, despite the fact that native English speakers in the "youth" group spend more time at English language websites than native English speakers in other age groups.

The results in table II also show 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. The coefficient falls in the 0.59-0.65 range for all three groups.

Table III shows regressions similar to that of table II. The only difference between the tables is that the regressions in table III include native English speakers and native French speakers who are bilingual. Table III shows that for the youngest age group, there is no significant difference between native English speakers and bilingual French speakers in terms of percentage use of English language websites (t=-0.69). In the case of the other two age groups, however, native French speakers who are bilingual use the Internet significantly less than native English speakers (t=-5.26 for the 21 to 45 age group and t=-3.22 for the 46 to 70 age group). These results suggest that English language websites are not that significant a barrier for French-speaking youth in Quebec.

IV.C Robustness of Results

In order to examine whether the results were robust, I conducted the following analyses:[18]

- I ran the same regressions reported in table II and table III 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 slightly smaller, but the results regarding the parameters of interest (the coefficients on Fbilingual and Language) are qualitatively unchanged.

- The estimates of the parameters of interest (coefficients on Fbilingual and Language) in tables II and III are qualitatively unchanged if we use the variables "Low Income" (a dummy variable that takes on the value one for income less than $25,000) and "High Income" (a dummy variable that takes on the value one for income greater than $100,000) instead of the variable "Income."

- The estimates of the parameters of interest (coefficients on Fbilingual and Language) in tables II and III are qualitatively unchanged if we eliminate the variable Age as an explanatory variable.

- It seems quite reasonable to assume that all right-hand-side variables in the regressions in table II and table III are exogenous in the short term, the period for which the empirical work is conducted. Hence, there is no need to use instrumental variable estimators in

these regressions. However, one might think about including Active Time as an explanatory variable in these regressions. In such a case, an argument can be made that active time itself is endogenous. Hence, I re-ran the regressions in table II and table III adding log(Active Time) to the set of explanatory variables. In such a case, it makes sense to instrument for log(Active Time), using the log(Active Time of Others) as the instrument. Again, I found that the key results on the Fblingual and Language coefficient estimates were quite robust to this alternative specification.

Hence, the main results regarding Fblingual and Language are quite robust to alternative specifications.

V. Conclusion

The results in section IV suggest that the difference between French and English speakers in the use of the Internet in English is relatively small for the younger generation. 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 that 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.

To the extent, however, 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. In such a case, as the youngest native French-speaking users age, they will continue to use the Internet intensively in English.

If there is indeed a cohort effect, there are potentially important implications for international trade. Given that common language typically leads to increased bilateral trade among countries,

these results (the significant use of English by non-native English speakers on the Internet) suggest that the Internet may lead to an increase in international trade. This is likely especially true in the case of services, the sector with the largest potential for trade growth. Currently, according to Freund and Weinhold [2002], while services account for 60 percent of production worldwide, they only account for 20 percent of international trade. Internet technology may provide the framework for substantially increasing international trade in services, especially if trade via the Internet uses a common language.

References

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, 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, "The Internet and International Trade in Services," American Economics Review, Papers and Proceedings, 92: 236-240, 2002.

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

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, "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, 1997.

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, 1999.

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

[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] Source: Global Reach at http://www.glreach.com/globstats/.

[4] Source: Global Reach at http://www.glreach.com/globstats/.

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

[7] 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.

[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.

[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, and 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.

[10] 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.

[11] Each user in the household had a different identification number, so that despite the multiple users in each household, the data are at the individual level.

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

[13] 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.

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

[15] See Grefenstette [1995] for an overview of language identification software programs.

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

[17] I am extremely grateful to Gregory Grefenstette and Clairvoyance.com for conducting the analysis on the 100,000 URL pages.

[18] These results are available from the author upon request.

Table I. 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	0	1
College_grad	0.43	0.49	0	1
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	0	1
College_grad	0.31	0.47	0	1
Kids	0.60	0.49	0	1
Active Time (hours)	9.86	15.93	0.08	114.47
Pereng	0.88	0.18	0	1

Table II: Weighted Least Squared Regression Results: Full Sample

Dependent Variable is $\log[p_i/(1-p_i)]$. These regressions are done at the level of the individual by age group. The t-statistics 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 III: Weighted Least Squared Regression Results: Native English speakers and Bilingual French Speakers Only

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

	Ages 10 to 20	Ages 21 to 45	Ages 46 to 70
Constant	-1.70 (2.04)	1.58 (1.66)	3.46 (1.66)
Age	0.14 (1.67)	0.019 (1.37)	-0.0038 (-0.11)
Female	-0.67 (-2.02)	-0.70 (-4.04)	-0.88 (-2.67)
Language	-0.22 (-0.69)	-1.11 (-5.26)	-1.12 (-3.22)
Size	0. 28 (2.13)	-0.14 (-1.23)	0.048 (0.28)
Income	-0.23 (-2.52)	0.070 (1.05)	-0.248(-2.60)
Kids	1.51 (3.08)	0.24 (1. 52)	-0.69 (-1.31)
HS_grad		0.14 (0.47)	0.28 (0.53)
College_grad		0.60 (1.24)	0.024 (0.28)
# of Obs.	57	95	48
R^2	0.41	0.39	0.30
Adj R^2	0.34	0.33	0.16

Figure I: Percent of Time Spent at English Language Web Sites

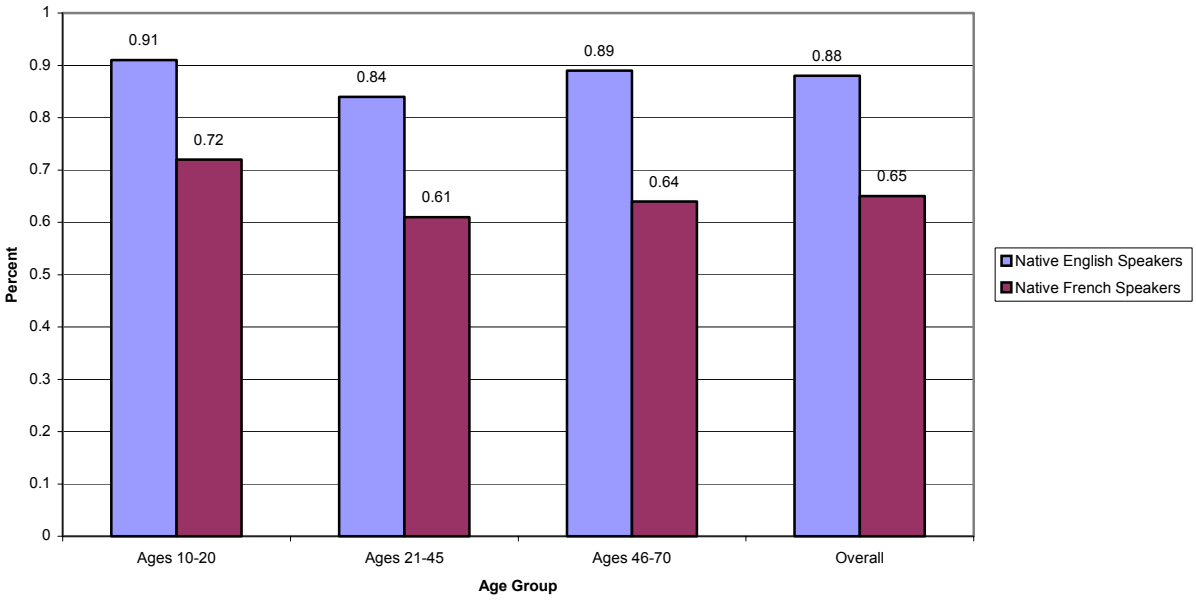

Figure II: Average Monthly Internet Use