DISCUSSION PAPER SERIES

IZA DP No. 1246

Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages

Barry R. Chiswick
Paul W. Miller

August 2004

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages

Barry R. Chiswick
University of Illinois at Chicago
and IZA Bonn

Paul W. Miller
University of Western Australia

Discussion Paper No. 1246
August 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 1246
August 2004

ABSTRACT

Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages[*]

This paper develops a scalar or quantitative measure of the “distance” between English and a myriad of other (non-native American) languages. This measure is based on the difficulty Americans have learning other languages. The linguistic distance measure is then used in an analysis of the determinants of English language proficiency among adult immigrants in the United States and Canada. It is shown that, when other determinants of English language proficiency are the same, the greater the measure of linguistic distance, the poorer is the respondent’s English language proficiency. This measure can be used in research, evaluation and practitioner analyses, and for diagnostic purposes regarding linguistic minorities in English-speaking countries. The methodology can also be applied to develop linguistic distance measures for other languages.

JEL Classification: I29, J15, J24

Keywords: immigrants, language proficiency, linguistic distance

Corresponding author:

Barry R. Chiswick
Department of Economics (M/C 144)
University of Illinois at Chicago
601 South Morgan Street (Room 2103 UH)
Chicago, IL 60607-7121
USA
Email: brchis@uic.edu

[*] Chiswick acknowledges the research support of the Institute of Government and Public Affairs, University of Illinois. Miller acknowledges the financial support of the Australian Research Council.

I. Introduction

This paper is concerned with the issue of “linguistic distance,” that is, the extent to which languages differ from each other. Although the concept is well known among linguists, the prevailing view is that it cannot be measured. That is, no scalar measure can be developed for linguistic distance.

In section II this paper discusses the concept of linguistic distance. Section III presents and discusses a scalar measure of the distance of other languages from English, based on the ease or difficulty Americans have in learning these other languages. This is followed (Section IV) by an application of this measure of linguistic distance to understanding the determinants of English language proficiency among adult immigrants from non-English speaking origins in the United States and Canada.[1] The paper closes (Section V) with a summary and conclusion.

II. Linguistic Distance

Studies of immigrant adjustment to the language of the host or destination country indicate that this adjustment differs significantly and substantially by country of origin, even after controlling statistically for the immigrant’s personal (socioeconomic and demographic) characteristics.[2] That is, immigrants from some countries of origin appear to be less proficient in the dominant language of the destination than do other immigrants, even when other measured variables are held constant. To some extent this may be due to different incentives for investing in destination language skills, such as the likelihood of temporary or permanent return migration, the availability of access to language training programs in the destination, or access to the destination language in the origin prior to migration. It would be expected, for example, that destination language skills would be greater among the foreign born if they did not expect to return to their origin, if they had access to destination language training in the destination, and if they were exposed to the destination language in schools, in the media or in the marketplace in the origin prior to migration.

Another reason why immigrant groups differ in their proficiency may be differences in the "distance" between the various immigrant languages and the destination language.[3] If English is linguistically "closer" to Western European languages (such as French and German) than it is to East Asian languages (such as Korean and Japanese), it would be expected that Western European immigrants in the U.S., UK, Canada and Australia would attain a higher level of proficiency in English, and would attain any given level of proficiency sooner, than immigrants from East Asia (see, for example, Corder, 1981, pp. 95-102).

Languages are complex. They differ in vocabulary, grammar, written form, syntax and myriad other characteristics. This makes for difficulty in the construction of measures of linguistic distance. Even if one intuitively "knows" that English is closer to French than it is to Chinese, by how much is it closer? If the difference is "large", how large is "large"? (McCloskey 1998, pp. 104-106). While it is easy to rank French as closer to English than Chinese is to English, other rankings of closeness to English may be more difficult, such as between Arabic and Russian or between Chinese and Japanese.

Linguists have developed models of the origins of languages and these models are expressed as "language trees." "The main metaphor that is used to explain the historical relationship is that of the language *family* or *family tree.*" (Crystal 1987, p. 292 italics in original). Through a language tree one may, in principle, trace the evolution of languages although linguists differ in their construction of language trees. Through a language tree it is possible to "trace" modern English back to its origins, but there is no measure of how different modern English is from its predecessor languages (Old English), other branches on the same tree (modern German), or even from languages on other trees (Chinese). While language trees are useful, they may be a poor guide to the qualitative distance across languages, and do not provide a quantitative measure.

A knowledge of linguistic distance may be invaluable for understanding differences across groups in the acquisition of destination language skills by adult and child immigrants, among participants in language training programs (such as, English as a second language or English for special purposes in the United States or abroad), or the linguistic issues facing indigenous linguistic minorities (e.g., indigenous language speaking peoples in Africa or Latin America), and the complexity of adaptation in multi-lingual societies (e.g., India and New Guinea).

Crystal (1987, p. 371) in The Cambridge Encyclopedia of Language writes regarding linguistic distance: "The structural closeness of languages to each other has often been thought to be an important factor in FLL (foreign language learning). If the L2 [the foreign language] is structurally similar to the L1 [the original language], it is claimed, learning should be easier than in cases where the L2 is very different. However, it is not possible to correlate linguistic difference and learning difficulty in any straightforward way, and even the basic task of quantifying linguistic difference proves to be highly complex, because of the many variables involved."[4] The many variables being the myriad characteristics that makes up the structure of languages.

It has been shown that “linguistic distance” affects the choice of destination among immigrants, and the language they adopt in multilingual destinations. For example, Chiswick and Miller (1994) show that immigrants to Canada are more likely to settle in Quebec if they came from a Romance language country rather than from a country with another mother tongue. Moreover, among immigrants in Quebec those from Romance language countries are more likely to become French-speakers while those from other (non-English) linguistic origins are more likely to become English language speakers.

Beenstock, Chiswick and Repetto (2001) show that among Jewish immigrants in Israel, those whose origin language was Arabic are the most proficient in Hebrew, other variables being the same. They suggest that this is due to the short linguistic distance between Hebrew and Arabic. Hebrew and Arabic, along with Amharic, are part of the Semitic branch of the Hamito-Semitic (Afro-Asiatic) family (Crystal 1987 and Grimes and Grimes, 1993). Among the languages included in the Israel analysis, Arabic is the closest to Hebrew. (The data were from the 1970s and there were negligible numbers of Ethiopian Jews in Israel at that time). A measure of linguistic distance from Hebrew comparable to the measure reported here for English has not yet been developed.

This paper reports a scalar or quantitative measure of the distance between English and a set of other languages. The value of this scalar measure of “linguistic distance” is demonstrated through an analysis of the determinants of English language proficiency among immigrants in two predominantly English-speaking immigrant receiving countries, the United States and Canada. The paper ends with a summary and conclusion.

III. Measuring Linguistic Distance

The quest among linguists for a scalar measure of linguistic distance has been in vain. There is no yardstick for measuring distances between or among languages, as there is for the geographic distance between countries (e.g., miles). This arises because of the complexity of languages, which differ by vocabulary, grammar, syntax, written form, etc. The distance between two languages may also depend on whether it is in the written or spoken form. For example, the written form of Chinese does not vary among the regions of China, but the spoken languages differ sharply. Alternatively, two languages that may be close in the spoken form may differ more sharply in the written form (for example, if they use different alphabets, as in the case of German and Yiddish).

Perhaps the way to address the distance between languages is not through language trees which trace the evolution of languages, but by asking a simpler question: How difficult is it for individuals who know language A to learn languages B_1 through B_i, where there are *i* other languages. If it is more difficult to learn language B_1, than it is to learn language B_2, it can be said that language B_1 is more "distant" from A than language B_2.[5] Language B_3 may be as difficult to learn as is language B_1 for a language A speaker, but that does not mean that language B_3 is close to language B_1. Indeed, it may be further from B_1 than it is from A.

Alternatively, if the issue is the adjustment of immigrants speaking languages B_1 through B_i in the linguistic destination A, one would want to know how difficult it is for speakers of B_1 through B_i to learn language A.[6] The U.S. Department of State, School of Language Studies teaches English-speaking Americans a variety of languages spoken in all corners of the globe, other than Native-American (American Indian) languages.

Achievements in speaking proficiency in these languages are then measured at regular intervals. For the same number of weeks of instruction a lower score represents less language facility, and it is assumed that this means a greater distance between the language and English. On the basis of the assumption of linguistic symmetry, this provides a measure of the linguistic distance between English and a variety of other languages.

The paper by Hart-Gonzalez and Lindemann (1993) reports language scores for 43 languages for English-speaking Americans of average ability after set periods (16 weeks and 24 weeks) of foreign language training. These languages are reported in the stub of Table 1, with their matching Census of Population Public Use Microdata Sample (PUMS) language codes for the 1990 and 2000 Censuses reported in the "direct code" column. Using the Ethnologue Language Family Index published by Grimes and Grimes (1993), the right-most column indicates the linguistic score for that language after 24 weeks of instruction based on the Hart-Gonzalez and Lindemann (1993) report. The range is from a low score (harder to learn) of 1.00 for Japanese to a high score (easier to learn) of 3.00 for Afrikaans, Norwegian and Swedish. The score for French is 2.50 and for Mandarin 1.50. These scores suggest a ranking of linguistic distance from English among these languages: Japanese being the most distant, followed by Mandarin, then French and then Afrikaans, Norwegian and Swedish as the least distant.

The data on language scores is extended to a much longer list of languages in the column labeled "close codes" (Grimes and Grimes, 1993). To as great an extent as possible, languages (other than Native American languages) not on the original list were matched with the assistance of a linguist for linguistic "closeness" to languages on the list.[7] Thus, Frisian (census code 612) is matched to Dutch (census code 610) which has a linguistic score of 2.75, and Icelandic and Farolse (census codes 617 and 618, respectively) are matched to Norwegian with a linguistic score of 3.00.

Language scores are reported in Table 1 for a wide range of languages that are spoken by foreign born and native born segments of the population in the United States. These scores can then be used to do statistical analyses of language issues.

IV. Application of the Measure of Linguistic Distance

This section reports the application of the measure of linguistic score in Table 1 to the analysis of proficiency in English among immigrants in the United States and Canada.

Using ordinary least squares regression analysis (OLS), Table 2 reports the partial effects of "linguistic distance" on the English language proficiency of foreign-born adult male and female immigrants in the United States from non-English speaking countries, using data from the 1990 Census of Population. The linguistic distance (LD) is measured in this analysis as the inverse of the linguistic score (LS) in Table 1, that is, LD = 1/LS. The "other variables held constant" include years of schooling, age and its square, duration in the U.S. and it square, marital status, a minority language concentration measure in the region of residence specific to the respondent's minority language, urban/rural residence and a south/non south region variable. Other variables the same, LD is a highly statistically significant variable for both men and women. Going from Swedish to Japanese (LD = 0.33 to LD = 1.0) reduces the probability of being proficient in English by about 17 percentage points (0.26 x 0.67 = 0.174), or the equivalent effect of about 5.4 years of additional schooling. The effect is larger (0.214) when the geographic distance (measured in miles) from the origin to the United States is also held constant. The partial regression coefficient and the t-ratio for the linguistic distance effect diminishes sharply (and disappears for men, but not for women) when country of origin is held constant through a set of dichotomous variables. This arises in large part because of the close correspondence of language and country—Korean is spoken in Korea, Italian in Italy, etc.

The linguistic distance measure was also applied to an analysis of English or French language proficiency among adult male immigrants in Canada from non-English speaking countries (Table 3).[8] Other variables the same, the greater the linguistic distance, the less likely is the immigrant to speak English, or if the immigrant speaks English, the less likely he is to speak English at home. At a duration in Canada of 5 years, only one quarter (25 percent) of immigrants with the greatest linguistic distance (LS = 1.0, Korean and Japanese) can carry on a conversation in English or French, in contrast to 5 percent for those with the smallest linguistic distance (LS = 3.0 Afrikaans, Swedish, Norwegian).

Even after 15 years in Canada, the ability to carry on a conversation in English or French varies by linguistic distance. Fully 10 percent of those with the greatest origin language linguistic distance cannot do so, compared to only 1 percent for those with the smallest distance. By 15 years in Canada only 5 percent of those with the greatest linguistic distance in their origin language usually speak English or French at home, in contrast to 58 percent for those with the shortest origin language distance. Thus, the linguistic patterns of immigrants in Canada, even after living there for 15 years, are influenced strongly by the distance between their origin language and English.

V. Summary and Conclusion

This paper develops and tests a scalar or quantitative measure of “linguistic distance.” Although linguists are familiar with the concept of the distance among the myriad characteristics of languages, the prevailing view is that it cannot be measured or quantified. This paper develops and tests such a measure.

The measure developed here is based on the ability of Americans to learn a variety of languages in fixed periods of time. The lower the scores on a standardized proficiency test, the greater is the distance between these languages and English. With the aid of a linguist, scores are inferred for languages for which a direct measure does not exist.

The measure of linguistic distance was then used in analyses of the English language proficiency of adult immigrants in the United States and Canada from non-English language origins, using census micro data. It is found empirically that the greater the distance between an immigrant’s origin language and English, the lower is the level of the immigrant’s English language proficiency, when other relevant variables are the same.

The measure of linguistic distance developed here can be used for other purposes. It can, for example, be used for research, evaluation, planning and diagnostic analyses for understanding the determinants of English language proficiency, in general or for specific purposes, among non-English speaking individuals, whether they are immigrants, non-English speaking linguistic minorities or learning in their country of origin.

The measure may also be useful for explaining patterns of international migration (i.e., choice of destination among immigrants), language adopted in multi-lingual destinations, and patterns of flows of tourists.[9] The measure can also be applied to other forms of analysis. Hutchinson (2002), for example, uses the linguistic distance measure developed for this paper in an analysis of international trade. He finds that, holding other relevant variables constant, a greater linguistic distance between the U.S. and other countries reduces both imports from and exports to the United States.

The methodology used here can, in principle, be developed for languages other than English. Thus, it would be possible to develop scalar measures of linguistic distance for other languages. This can permit the development of a full range of measures of linguistic distance.

Table 1

Index of Difficulty of Learning a Foreign Language (Language Scores) and Codes for Languages Reported in the U.S. Census

Language	Direct Codes 1990, 2000 Censuses	Close Codes 1990 Census	Changes for 2000 Census	Language Score
Afrikaans	611			3.00
Danish	615			2.25
Dutch	610	612		2.75
French	620	621,622,623,624		2.50
German	607	608,609,613		2.25
Italian	619			2.50
Norwegian	616	617,618		3.00
Portuguese	629	630		2.50
Rumanian	631	632		3.00
Spanish	625	626, 627		2.25
Swedish	614			3.00
Indonesian	732	730-731, 733-737		2.00
Malay	739			2.75
Swahili	791	792		2.75
Amharic	780			2.00
Bengali	664			1.75
Bulgarian	647	648		2.00
Burmese	717			1.75
Czech	642			2.00
Dari	660			2.00
Farsi	656	657, 658, 659, 661		2.00
Finnish	679	680		2.00
Greek	637			1.75
Hebrew	778			2.00

Table 1 Continued

Hindi	663	662, 665-669, 678	Add 671	1.75
Hungarian	682			2.00
Lao	720			1.50
Cambodian	726			2.00
Mongolian	694	695, 716		2.00
Nepali	674			1.75
Polish	645	644, 646		2.00
Russian	639	640, 641		2.25
Serbo-Croatian	649-651	652		2.00
Sinhala	677			1.75
Tagalog	742	740, 741, 743-749		2.00
Thai	720	717, 718, 719	Add 725	2.00
Turkish	691	689, 690, 692, 693		2.00
Vietnamese	728	729		1.50
Arabic	777	779		1.50
Mandarin	712	713, 714, 715		1.50
Japanese	723	725	Delete 725	1.00
Korean	724			1.00
Cantonese	708	709, 710, 711, 721, 722		1.25

Note: Language Codes in this table are from *1990 United States Census of Population and Housing, Technical Documentation* and from *2000 United States Census of Population and Housing, Technical Documentation.* There are minor differences in the language codes in the 1990 and 2000 Censuses. These differences are indicated in column (3). Column (4) is the language score for the direct codes.

Source of Matching Codes: (a) Joseph E. Grimes and Barbara F. Grimes, *Ethnologue: Languages of the World*, Summer Institute of Linguistics Inc., Dallas, Texas, 13th edition, 1993.
(b) Adam Makkai, Professor of Linguistics, Department of English, University of Illinois at Chicago.

Source of Language Score: Lucinda Hart-Gonzalez and Stephanie Lindemann, "Expected Achievement in Speaking Proficiency, 1993", School of Language Studies, Foreign Services Institute, Department of State, April 15, 1993.

Table 2

Partial Effect of Linguistic Distance on the English Language Proficiency of Foreign-Born Adults from Non-English Speaking Countries, 1990 U.S. Census[(a)]

		Males	Females
1)	Other variable held constant[(b)]	-0.256 (-44.91)	-0.263 (-51.95)
2)	Other variables and distance of foreign country from the U.S. in miles, and its square	-0.319 (-53.34)	-0.320 (-60.10)
3)	Other variables, distance in miles and its square, and country fixed effects	0.007 (0.56)	-0.063 (-4.39)

NOTE: Sample Size 237,770 for males and 243,496 for females.

[(a)]The measure of Linguistic Distance (LD) is the inverse of the Linguistic Score (LS) in Table 1. That is LD = 1/LS from Table 1. The dependent variable is unity if the respondent speaks only English at home or, if another language is spoken, English is spoken "very well" or "well". It is zero for those who speak English "not well" or "not at all." The foreign-born excludes those born in the English-speaking developed countries (U.K., Ireland, Canada, Australia and New Zealand). Adults are persons age 25 to 64 in 1990. Where only English is spoken at home, and hence a non-English language is not reported, LS is the mean value of the linguistic score measure for individuals reporting a foreign language from their birthplace group.

[(b)]Other variables held constant include years of schooling, age (and its square), duration of residence in the United States (and its square), marital status, an index of the extent to which their origin language is spoken in their state of residence, and variables for urban/rural and south/ non-south residence.

[(c)] Country fixed effects represented by 16 country/region of birth dichotomous variables.

Note: t-ratios are in parentheses.

Source: Chiswick and Miller (1998, Tables 2 and 6).

Table 3

Predicted Distributions across Language Categories by Linguistic Score and Duration of Residence, Foreign-Born Adult Males from Non-English Speaking Countries, 1991 Census of Canada[a]

Linguistic	After 5 years in Canada			After 15 years in Canada		
Score[b]	E1	E2	E3	E1	E2	E3
1.0	24.54	73.88	1.58	10.05	85.35	4.60
2.0	8.00	73.75	18.25	2.32	60.18	37.51
3.0	4.57	61.18	34.24	1.09	41.05	57.86

NOTE: Sample Size 32, 168.

[a] Predicted values from a multinomical logit model. Adults are age 25 to 64 in 1990, Foreign-born exclude those born in the U.S., U.K. and Ireland.
E1 = Cannot carry on a conversation in English or French.
E2 = Can carry on a conversation in English of French, but usually speak another language at home.
E3 = Can carry on a conversation in English or French and usually speak one of these languages at home.

[b] Language scores range from 1.0 (Japanese and Korean) to 3.0 (Afrikaans, Norwegian and Swedish).

Source: Chiswick and Miller (2001, Table 5).

REFERENCES

Beenstock, Michael (1996). "The Acquisition of Language Skills by Immigrants: The Case of Hebrew in Israel," International Migration, 3, pp. 3-30.

Beenstock, Michael, Barry R. Chiswick and Gaston L. Repetto (2001). "The Effect of Linguistic Distance and Country of Origin on Immigrant Language Skills: Application to Israel," International Migration, 39(3), pp. 33-60.

Chiswick, Barry R. (1998). "Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel," Journal of Population Economics, 11(2), May, pp. 253-271.

Chiswick, Barry R. and Paul W. Miller (1994). "Language Choice Among Immigrants in a Multi-Lingual Destination," Journal of Population Economics, 7, pp. 119-131.

Chiswick, Barry R. and Paul W. Miller (1995). "The Endogeneity Between Language and Earnings: International Analyses," Journal of Labor Economics, 13, pp. 246-288.

Chiswick, Barry R. and Paul W. Miller (1996). "Ethnic Networks and Language Proficiency Among Immigrants," Journal of Population Economics, 9, pp. 19-35.

Chiswick, Barry R. and Paul W. Miller (1998). "English Language Fluency Among Immigrants in the United States," Research in Labor Economics, 17, pp. 151-200.

Chiswick, Barry R. and Paul W. Miller (2001). "A Model of Destination Language Acquisition: Application to Male Immigrants in Canada," Demography, 38(3), August, pp. 391-409.

Corder, S. Pit (1981) Error Analysis and Interlanguage, Oxford: Oxford University Press.

Crystal, David (1987). The Cambridge Encyclopedia of Language, Cambridge: Cambridge University Press.

Dustmann, Christian (1997). "The Effects of Education, Parental Background and Ethnic Concentration on Language," Quarterly Review of Economics and Finance, 37, Special Issue, pp. 245-262.

Dustmann, Christian and Francesca Fabbri (2003). "Language Proficiency and Labour Market Performance of Immigrants in the UK," Economic Journal, July, pp. 695-717.

Ellis, Rod (1994). The Study of Second Language Acquisition, Oxford: Oxford University Press.

Grenier, Giles and Francois Vaillancourt (1983). "An Economic Perspective on Learning a Second Language," Journal of Multilingual and Multicultural Development, 4, pp. 471-483.

Grimes, Joseph E. and Barbara F. Grimes, (1993). Ethnologue: Languages of the World, Thirteenth Edition, Dallas: Summer Institute of Linguistics, Inc.

Hart-Gonzalez, Lucinda and Stephanie Lindemann, (1993). "Expected Achievement in Speaking Proficiency, 1993," School of Language Studies, Foreign Services Institute, Department of State, mimeo.

Hutchinson, William K. (2002). "Linguistic Distance as a Determinant of Bilateral Trade," Department of Economics, Vanderbilt University, Xerox.

McCloskey, Deirdre (1998). The Rhetoric of Economics, Madison: University of Wisconsin Press, 2nd edition.

Shields, Michael A. and Stephen Wheatley Price (2002). "The English Language Fluency and Occupational Success of Ethnic Minority Immigrant Men Living in English Metropolitan Areas," Journal of Population Economics, 15(1), January, pp. 137-160.

U.S. Bureau of the Census (1993). 1990 United States Census of Population and Housing, Technical Documentation, Washington, D.C.

U.S. Bureau of the Census (2003). 2000 United States Census of Population and Housing, Technical Documentation, Washington, D.C.

ENDNOTES

[1] For an analysis of the determinants of second language acquisition from an economist's perspective see Chiswick and Miller (1998) and from a linguist's perspective see Ellis (1994).

[2] For example, these studies have been conducted for the United States (Chiswick and Miller, 1998), Australia (Chiswick and Miller, 1995, 1996), Canada (Grenier and Vaillancourt, 1983, Chiswick and Miller 2001), Germany (Dustmann 1997), Israel (Beenstock 1996, Chiswick 1998) and the United Kingdom (Shields and Wheatley Price 2002, Dustmann and Fabbri, 2003).

[3] The story in Genesis about the Tower of Babel emphasizes the difficulty of working cooperatively when there is a lack of communication among individuals based on differences in languages.

[4] In their study of the English language proficiency of immigrants in the UK, Shields and Wheatley Price (2002, p. 145) indicate that their theoretical model calls for a measure of linguistic distance of the immigrants' origin language from English, but they do not have a direct measure and they use country of birth dichotomous variable to reflect this and other origin-specific effects.

[5] Think of each language as having N dimensions, where the N dimensions represent the various aspects of language (Crystal, 1987, p.371). Then each language can be thought of as being represented by a point in N-dimensional space, and could be described by a vector $(a^1, a^2, ..., a^N)$, $(b_1^1, b_1^2, ..., b_1^N)$, etc, where a^n is the amount of the nth dimension that characterizes language A, and b_i^n is the amount of the nth dimension that characterizes language B_i, etc. The distance between any two languages is given by the Euclidean distance function. The measure of linguistic distance proposed in this paper can be thought of as a proxy for the Euclidean distance between language A and the various languages B_1 though B_i.

[6] While the linguistic difference between A and, say. B_i, is a given magnitude, the impact of that measure of distance on proficiency in language B_i, among language A speakers may differ from the impact of the distance on language B_i speakers learning language A.

[7] We are indebted to Adam Makkai, Professor of Linguistics, Department of English, University of Illinois at Chicago for helping us with this coding.

[8] Those born in France cannot be separately identified in the Canadian Census due to the small number of immigrants from France.

[9] Other variables the same, tourist flows would be expected to be greater the smaller the linguistic distance between the languages of the origin and tourist destination.