

Galor, Oded; Özak, Ömer; Sarid, Assaf

Working Paper

Linguistic Traits and Human Capital Formation

GLO Discussion Paper, No. 570

Provided in Cooperation with:
Global Labor Organization (GLO)

Suggested Citation: Galor, Oded; Özak, Ömer; Sarid, Assaf (2020) : Linguistic Traits and Human Capital Formation, GLO Discussion Paper, No. 570, Global Labor Organization (GLO), Essen

This Version is available at:
<http://hdl.handle.net/10419/218948>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Linguistic Traits and Human Capital Formation*

Oded Galor[†], Ömer Özak[‡] and Assaf Sarid[§]

December 11, 2019

Abstract

This research establishes empirically that existing cross-language variations in the structure of the future tense and the presence of grammatical gender affected human capital accumulation. Exploiting variations in the dominant languages among migrants from the same countries of origin, the study explores the impact of these traits on the educational attainment of second generation migrants in the US. The results suggest that college attendance among individuals with identical ancestry is (i) higher if the dominating language at home has a periphrastic future tense, and (ii) lower for women exposed predominantly to sex-based grammatical gender.

Keywords: Comparative Development, Human Capital, Education, Language Structure, Future Tense, Grammatical Gender, Cultural Evolution, Gender Bias, Long-term Orientation

JEL Classification: I20, J16, Z10, Z13

*The authors are grateful to Joan Bybee, Justin Cook, Delia Furtado, Paola Giuliano, Luigi Guiso, Klaus Desmet, Stelios Michalopoulos, and David Weil, as well as conference participants at the “Annual Meetings of the American Economic Association”, 2016; “Deep Rooted Factors in Comparative Development”, 2016, Zeuthen Lectures, Copenhagen, 2016, and seminar participants at Ben-Gurion, Brown, Clark, Haifa, Southern Methodist, Tel-Aviv, and UC - Merced, for useful comments and discussions.

[†]Department of Economics, Brown University; NBER, CEPR, IZA, CES-Ifo. E-mail: Oded_Galor@brown.edu

[‡]Department of Economics, Southern Methodist University. E-mail: ozak@smu.edu

[§]Department of Economics, University of Haifa. Email: asarid@econ.haifa.ac.il

1 Introduction

The origins of the vast inequality in the wealth of nations have recently been attributed to the persistent effect of an uneven distribution of pre-industrial geographical, cultural, institutional and human characteristics across the globe.¹ In light of the coevolution of cultural and linguistic characteristics in the course of human history, the evolution of language has conceivably reinforced the persistent effect of cultural factors on the process of development.

This research explores some of the most fundamental and intriguing mysteries about the impact of the coevolution of linguistic and cultural traits on the development process:² Has the coevolution of linguistic and cultural traits contributed to the persistence of cultural characteristics and their lasting effect on economic prosperity? Have language structures merely reflected existing cultural traits or have they influenced human behavior and values and contributed directly to the development process?

In view of the pivotal role of language in the transmission of knowledge and values, language structures have plausibly affected the diffusion of cultural values and thus human behavior across members of society, reinforcing existing cultural traits and their intergenerational transmission. In particular, in light of the communication function of language, emerging language structures have conceivably facilitated efficient communication across individuals, while enhancing the transmission of cultural values. The forces of natural selection across language structures may have therefore generated an evolutionary advantage to those structures that reflected dominating cultural traits. Furthermore, considering the pivotal role of language as a coordination device across members of society, the evolution of language structures necessitated and reflected the adoption of linguistic mutations by society as a whole. Unlike the feasibility of a unilateral deviation by individuals from existing cultural norms, the diffusion of unilateral linguistic innovations is unlikely and language structures would therefore tend to be more persistent than cultural traits. Thus, inevitably, cultural traits reflected in language structures would be expected to be more persistent across time and space. Moreover, it is not inconceivable that language structures per se may have directly influenced individuals' mindsets and thus human behavior, beyond the non-linguistic transmission channel of culture.³

In particular, a society characterized by distinct gender roles and consequently by the existence of gender bias, grammatical gender that could have fortified the existing social structure and cultural norms may have emerged and persisted over time. Similarly, in societies characterized by long-term orientation, a structure of the future tense that could have reinforced the efficiency of future oriented behavior may have emerged and persisted over time.

The research suggests that while linguistic traits have been largely a reflection of past human

¹Gallup et al. (1999), Guiso et al. (2004, 2006), Tabellini (2010), Acemoglu et al. (2001), Glaeser et al. (2004), and Ashraf and Galor (2013).

²Existing economic research predominantly views languages as an identifier of cultural and ethnic groups. Linguistic fractionalization as well as linguistic distance have been extensively used as a proxy for ethnic fractionalization and cultural distance in the exploration of the effect ethnic diversity on economic growth and the impact of cultural distance on the diffusion of development (Easterly and Levine, 1997; Fearon, 2003; Alesina et al., 2003; Alesina and Ferrara, 2005; Desmet et al., 2012; Harutyunyan and Özak, 2016).

³The Oxford English Dictionary defines mindset as “[a]n established set of attitudes, esp. regarded as typical of a particular group’s social or cultural values; the outlook, philosophy, or values of a person;”... “an incident of a person’s *Weltanschauung* or philosophy of life”.

experience, and in particular ancestral cultural traits, they have played a pivotal role in the persistent effect of cultural characteristics on comparative economic development. Moreover, the evidence suggests that there exists a direct and independent effect of language structures on human behavior and contemporary economic outcomes.

The empirical analysis examines the effects of language structures on contemporary economic outcomes, conceivably via their potential impact on the persistence of ancestral cultural traits as well as on individual behavior. Several strategies are employed in order to surmount the significant hurdles in the identification of the effect of language-embodied cultural traits on human behavior, while isolating this effect from the persistent effect of cultural traits via non-linguistic channels. Following the epidemiological approach for the identification of the persistent effects of cultural traits on human behavior and economic outcomes (Giuliano, 2007; Fernandez and Fogli, 2009), the research focuses on the behavior of second-generation migrants who share the same country of birth, removing concerns about the role of geographical, institutional and cultural characteristics in this country.

Nevertheless, the traditional epidemiological approach cannot fully distinguish between the persistent effect of observed cultural characteristics and omitted ancestral characteristics at the parental countries of origin. In particular, since the basic premise of the epidemiological approach is that second-generation migrants from the same parental countries of origin share the cultural heritage of those countries, the traditional epidemiological approach cannot account for parental countries of origin fixed-effects. In contrast, in light of the fact that second-generation migrants that have identical parental countries of origin may speak different languages, one can isolate the effect of language-embodied cultural traits on human behavior, by accounting for parental countries of origin fixed-effects (i.e., common ancestral factors such as geographical, institutional and cultural characteristics that may affect individual behavior). Hence, the analysis exploits variations in language structures across individuals that are originated from the same ancestral homelands in order to identify the effect of language-embodied cultural traits on human behavior.

The analysis focuses on the effect of (i) the presence periphrastic future tense and its association with long-term orientation on educational attainment, and (ii) the presence of sex-based grammatical gender and its association with gender bias on female educational attainment. The analysis establishes a beneficial effect of speaking a language with periphrastic future tense (associated with long-term orientation) on college attendance and an adverse effect of speaking a language with sex-based grammatical gender (associated with gender-bias) on female college attendance.

The empirical methodology advanced in the course of this research augments the epidemiological approach and advances a methodology that permits the isolation the effect of cultural traits that are language-embodied on human behavior from the persistent effects of culture via non-linguistic channels. This advancement overcomes some of the limitations of the existing studies about the association between language structures and economic outcomes (Chen, 2013; Roberts et al., 2015).

Languages differ in the structure of their future tense. In particular, linguists distinguish between languages that are characterized by an *inflectional* versus *periphrastic* future tense (Dahl, 1985, 2000; Dahl and Velupillai, 2013). Inflectional future tense is associated with verbs that display morphological variation (i.e., a change in the verb form that is associated with the future tense). In contrast,

periphrastic future tense is characterized by roundabout or discursive phrases (e.g., expressions such as ‘will’, ‘shall’, ‘want to’, ‘going to’ in the English language) (Bybee and Pagliuca, 1987; Bybee and Dahl, 1989; Bybee et al., 1994).

As argued by Bybee and Dahl (1989), unlike inflectional future tense, periphrastic future tense are formed by terms that express a desire, an intention, an obligation, a commitment as well as a movement towards a goal. In particular, in the case of English, “*shall* has developed from a main verb meaning ‘to owe’, *will* from a main verb meaning ‘to want’, and the source of *be going to* is still transparent” (Bybee and Dahl, 1989, p.90).

Moreover, “intention and prediction are most commonly expressed by the periphrastic future, while the synthetic one is more common in generic statements, concessives, and suppositions” (Bybee et al., 1994, p.235). Inflectional futures “also appear systematically (often obligatorily) in sentences which express clear predictions about the future (which are independent of human intentions and planning), whereas less grammaticalized constructions [i.e., periphrastic] often tend to be predominantly used in talk of plans and intentions—a fact which is explainable from the diachronic sources of future tenses” (Dahl and Velupillai, 2013, p.270). Thus, periphrastic future tense captures long-term oriented intentions.

Languages differ in the existence and the form of grammatical gender. In particular, languages that are characterized by sex-based grammatical gender classify nouns according to biological gender. The presence of sex-based grammatical gender induces speakers to highlight gender distinctions even in situations in which gender may not play an intrinsic role. Moreover, linguists have argued that the presence of “masculine generics” in some languages (e.g., the use of the noun “adam” to describe “man” as well as “human being” in the Hebrew and Turkish languages) has reinforced gender biases in the course of human history, equating “maleness and humanness” (Stahlberg et al., 2007, p.169). Thus, perhaps not surprisingly, linguists as well as other scholars have persistently argued that gender biases have been reinforced by languages characterized by sex-based grammatical gender systems Lakoff (1973); Eckert and McConnell-Ginet (2003); Stahlberg et al. (2007).

2 Language Structures & Education

This section explores the potential effect of the linguistic channel of cultural transmission on contemporary behavior. The analysis focuses on the effect of (i) the presence periphrastic future tense and its association with long-term orientation on educational attainment, and (ii) the presence of sex-based grammatical gender and its association to gender bias on female educational attainment.

2.1 Identification Strategy

Several strategies are employed in order to surmount the significant hurdles in the identification of the effect of language-embodied cultural traits on human behavior, while isolating this effect from the persistent effect of cultural traits via non-linguistic channels.

Following the epidemiological approach for the identification of the persistent effects of cultural traits on human behavior and economic outcomes (Giuliano, 2007; Fernandez and Fogli, 2009; Galor

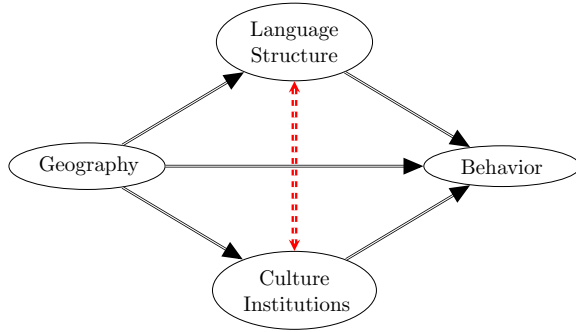


Figure 1: Language and Contemporary Behavior

and Özak, 2016), the research focuses on the behavior of second-generation migrants who share the same country of birth, removing concerns about the role of geographical, institutional and cultural characteristic in this country. Moreover, the analysis accounts for individual characteristics (e.g., age, gender, and marital status), which might affect individual behavior while being correlated with the language spoken by the individual. Finally, it accounts for year and locality fixed-effects, further removing the potentially confounding effects of the period as well as the local geography, culture, institutions, and socio-economic environment.

Nevertheless, the traditional epidemiological approach cannot fully distinguish between the persistent effect of observed cultural characteristics and omitted ancestral characteristics at the parental countries of origin. In particular, since the basic premise of the epidemiological approach is that second-generation migrants from the same parental countries of origin share the cultural heritage of those countries, the traditional epidemiological approach cannot account for parental countries of origin fixed-effects. In contrast, in light of the fact that second-generation migrants that have identical parental countries of origin may speak different languages, one can isolate the effect of language-embodied cultural traits on human behavior, by accounting for parental countries of origin fixed-effects (i.e., common ancestral factors such as geographical, institutional and cultural characteristics that may affect individual behavior).

Hence, the analysis exploits variations in language structures across individuals that are originated from the same ancestral homelands in order to identify the effect of language-embodied cultural traits on human behavior, while also accounting for confounding parental characteristics such as education and the level of proficiency in the local language.^{4,5}

⁴Data is taken from the US Census and American Community Survey for the years post-2000 based on IPUMS (Ruggles et al., 2015). Second-generation migrants include all US-born individuals with at least one foreign born parent. The data on second generation migrants include 165250 offsprings of parents who migrated to the United States from 138 different countries – 137 countries of origin of the mother and 136 countries of origin of the father; these individuals speak 62 different languages. The sample of second-generation migrants in the US is constrained to include only individuals over 24 years of age in order to ensure they are old enough to have attended college. As shown in the appendix, similar results are obtained if the age is constrained to be over 21 or 18.

⁵By focusing on second-generation migrants in the US Census and American Community Surveys (ACS), the analysis overcomes a potential concern due to ethnic attrition bias (Duncan and Trejo, 2016). In particular, previous analyses that have employed the US census or ACS to study the effects of culture using migrants, have focused on all US-born individuals and tried to identify migrants and their ancestry by using individual’s self-reported ancestry. Thus, these analyses have included all descendants of migrants that still identify with the country of origin of their ancestors. But, as Duncan and Trejo (2011, 2016), among others, have shown, individuals tend to self-identify differently depending on

2.2 Crop Return, Periphrastic Future Tense, and Long-Term Oriented Behavior of Second-Generation Migrants

This section explores the language-embodied effect of long-term orientation, as reflected in periphrastic future tense, on the long-term oriented behavior of its speakers as opposed to the persistent effect of cultural traits via non-linguistic channels.⁶ Given the data requirements for the identification strategy discussed in the previous section, the analysis focuses on the effect of periphrastic future tense on human capital accumulation of second-generation migrants in the US. In particular, it explores the effect of speaking a language with periphrastic future tense on the probability of college attendance of these second-generation migrants.

In order to analyze the effect of periphrastic future tense on college attendance, the following general specification is estimated via ordinary least squares (OLS):

$$\text{College}_{istlp} = \beta_0 + \beta_1 \text{Future}_{istlp} + \beta_2 \text{Return}_{istlp} + \sum_j \gamma_{0j} X_{istlpj} + \sum_{stpj} \gamma_{stpj} \delta_{stpj} + \epsilon_{istlp}, \quad (1)$$

where College_{istlp} indicates whether individual i in state s in period t who speaks language l with parental ancestry p has attended college or not, Future_{istlp} indicates the existence of periphrastic future tense in language l spoken by the individual, the Return_{istlp} is the pre-1500CE crop return in the homeland of language l spoken by the individual, $\{X_{istlpj}\}_j$ is a set of additional geographical characteristics of the homeland of the language spoken by the individual, $\{\delta_{stpj}\}_j$ is a set of fixed-effects that account jointly for individual characteristics j (sex, age, marital status), state s , year t , and parental ancestry p , and ϵ_{istlp} is the error term. Thus, the fixed-effects ensure that only individuals that are similar in their observable individual characteristics, their location and ancestry are compared to each other. The results on the origins of the periphrastic future tense presented in the historical analysis in section ?? suggest that future tense and pre-1500 crop return should have a positive effect on college attendance (i.e. $\beta_1 > 0$ and $\beta_2 > 0$).

Table 5 establishes the positive effect of speaking a language with periphrastic future tense on college attendance of its speakers. In particular, columns (1)-(3) show that individuals who speak a language with periphrastic future tense have 20 percentage points higher probability of attending college than individuals with similar observable characteristics living in the same state and interviewed the same year, who do not speak a language with periphrastic future tense, even after accounting geographical characteristics of the homeland of the language.

As mentioned in the identification strategy, one potential concern with the results of columns (1)-(3) is that the estimated effect of language also captures additional cultural elements due to the ancestry of the individual. In order to overcome this potential concern, column (4) additionally accounts for the parental country of origin. Thus, the estimated effect of periphrastic future tense in column (4) captures the effect of language that is not explained by other ancestral traits, and therefore isolates the effect of long-term orientation that is language-embodied from the persistent cultural effects of

their generation, their true ancestry, and their socio-economic background. Thus, using second-and-higher-generation migrants can bias the results due to misidentification of ancestry.

⁶The effect of long-term orientation on human capital accumulation via non-linguistic channels is studied by Galor and Özak (2016) and Figlio et al. (2016).

Table 1: Periphrastic Future Tense and College Education of Second Generation Migrants

	College Attendance							
	All					No ENG	NO SPA	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Periphrastic Future Tense	0.228*** (0.055)	0.223*** (0.052)	0.040*** (0.015)	0.038*** (0.012)	0.038** (0.015)	0.035*** (0.012)	0.062*** (0.023)	0.032** (0.015)
Mom's Education Level (HS+)				0.133*** (0.012)		0.135*** (0.011)	0.121*** (0.016)	0.124*** (0.013)
Dad's Education Level (HS+)				0.137*** (0.014)		0.139*** (0.014)	0.128*** (0.021)	0.130*** (0.014)
Mom's English Level					0.013*** (0.004)	0.014*** (0.002)	0.016*** (0.001)	0.006* (0.003)
Dad's English Level					0.002 (0.003)	0.007*** (0.002)	0.008*** (0.002)	0.003 (0.003)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parental Origin FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.06	0.09	0.13	0.17	0.13	0.17	0.18	0.18
R^2	0.06	0.14	0.18	0.22	0.18	0.22	0.23	0.24
Observations	130455	130455	130455	130455	130455	130455	74709	75664

Notes: Heteroskedasticity robust standard error estimates clustered at the parental countries of origin, language and state levels are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

long-term orientation via non-linguistic channels. The results suggest that speaking a language with periphrastic future tense increases the probability of attending college by 4.6 percentage points.

Additionally, Column (5) establishes that the pre-1500 crop return in a language's homeland has a positive effect on the accumulation of human capital of its speakers, even after accounting for all other ancestral characteristics of an individual and other geographical characteristics of the language's homeland. Column (6) provides supportive evidence to the view that periphrastic future tense reflected the cultural effect of crop return. In particular, it suggests that the effect of crop return is mediated by a language's periphrastic future tense. Thus, columns (5) and (6) support the view that the effect of periphrastic future tense partly captures the persistent effect of cultural traits that reflect crop return and thus long-term orientation.

There are various potential concerns with the results of Table 5. First, second-generation migrants in the US Census and ACS can only be identified for individuals who live with their parents. Although this is a representative sample of this subpopulation, which overcomes concerns due to ethnic attrition (see footnote 5) and allows for the control of parental characteristics in the analysis, it might potentially

Table 2: Periphrastic Future Tense and College Education of One-and-a-half Generation Migrants

	College Attendance							
	All						No ENG	NO SPA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Periphrastic Future Tense	0.228*** (0.062)	0.224*** (0.061)	0.065** (0.025)	0.078*** (0.018)	0.065*** (0.016)	0.073*** (0.017)	0.082*** (0.026)	0.056* (0.030)
Mom's Education Level (HS+)				0.129*** (0.013)		0.132*** (0.013)	0.112*** (0.020)	0.123*** (0.021)
Dad's Education Level (HS+)				0.130*** (0.012)		0.130*** (0.013)	0.123*** (0.023)	0.124*** (0.015)
Mom's English Level					0.015*** (0.003)	0.015*** (0.002)	0.014*** (0.002)	0.008** (0.003)
Dad's English Level					0.007** (0.003)	0.010*** (0.004)	0.007** (0.003)	0.020*** (0.006)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.06	0.10	0.14	0.19	0.16	0.19	0.21	0.18
R^2	0.06	0.15	0.19	0.29	0.26	0.29	0.31	0.31
Observations	513028	513028	513028	30104	30104	30104	19664	17187

Notes: Heteroskedasticity robust standard error estimates clustered at the country of origin, language and state levels are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

bias the results. Appendix A.8 explores the differences in observables between various samples of migrants. Reassuringly, it shows that only age and marital status differ between the full sample of second-generation migrants and the subsample that lives with their parents. Moreover, the sample of second-and-higher generation migrants, that has previously been employed in the literature, and which is subject to ethnic attrition bias, is more similar to the true third-and-higher-generation migrant sample.

In order to assess the potential bias due to the sample, Table A.1 replicates the basic results (without ancestry fixed-effects given the potential for ethnic attrition bias) for the sample of second-and-higher generation migrants. Additionally, Table A.2 replicates the analysis in Table 5 using the sample of one-and-a-half-generation migrants, i.e., migrants who were born in another country, but arrived to the US before age 5. The benefits of using this sample is that (i) it has similar properties for cultural analysis as second-generation migrants, and (ii) it overcomes the potential concerns due to both ethnic attrition and living arrangements.⁷ Reassuringly, the qualitative results remain unchanged

⁷The sample of the one-and-a-half-generation migrants includes 422081 individuals who migrated from 141 different countries to the United States when they were five years old or younger and speak 64 different languages. One-and-a-

Table 3: Sex-Based Grammatical Gender and Female College Education of Second Generation Migrants

	College Attendance							
	All						No ENG	NO SPA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Existence of Sex-Based Gender System	-0.201*** (0.047)	-0.181*** (0.044)	-0.013 (0.021)	-0.038** (0.018)	-0.009 (0.020)	-0.034* (0.018)	-0.064*** (0.018)	-0.036* (0.021)
Mom's Education Level (HS+)				0.123*** (0.010)		0.125*** (0.010)	0.112*** (0.015)	0.117*** (0.015)
Dad's Education Level (HS+)				0.127*** (0.015)		0.128*** (0.015)	0.123*** (0.018)	0.120*** (0.012)
Mom's English Level					0.014*** (0.005)	0.014*** (0.004)	0.017*** (0.002)	-0.000 (0.003)
Dad's English Level					0.003 (0.003)	0.008*** (0.002)	0.009*** (0.001)	0.009*** (0.003)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parental Origin FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.05	0.08	0.11	0.14	0.11	0.14	0.14	0.17
R^2	0.07	0.14	0.17	0.20	0.17	0.20	0.20	0.26
Observations	52734	52734	52734	52734	52734	52734	29903	27339

Notes: Heteroskedasticity robust standard error estimates clustered at the parental countries of origin, language and state levels are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

and suggest that speaking a language with periphrastic future tense increases the probability of college attendance by 5 percentage points, above and beyond the effect of other ancestral traits.

Second, individuals' education levels are potentially determined by the education level of their parents. Similarly, parents' command of the English language, which is the official language in the US, might potentially affect individual's education levels as well as the language spoken at home. Table 6 explores the effect of accounting for parents' education levels and their command of the English language. Additionally, the analysis accounts now for fixed-effects for both parents' countries of origin. Reassuringly, the results of Table 5 remain qualitatively unchanged. In particular, the effect of speaking a language with periphrastic future tense remains positive and significant. Additionally, parental education and English levels have a positive effect on their offspring's college attendance, suggesting that college educated parents who have a good command of English have a higher probability of having their children attend college. The estimates suggest that speaking a language with periphrastic future

half-generation migrants are similar to second-generation migrants, since they were not the ones who made the decision to migrate and grew up in the US, so that they received their K-12 education in the United States.

Table 4: Sex-Based Grammatical Gender and Female College Education of One-and-a-half Generation Migrants

	College Attendance							
	All						No ENG	NO SPA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Existence of Sex-Based Gender System	-0.238*** (0.067)	-0.233*** (0.061)	-0.069** (0.025)	-0.106*** (0.033)	-0.053 (0.032)	-0.096*** (0.034)	-0.139* (0.068)	-0.086* (0.043)
Mom's Education Level (HS+)				0.128*** (0.013)		0.130*** (0.014)	0.116*** (0.015)	0.129*** (0.017)
Dad's Education Level (HS+)				0.108*** (0.015)		0.107*** (0.014)	0.098*** (0.030)	0.094*** (0.020)
Mom's English Level					0.022*** (0.003)	0.022*** (0.002)	0.021*** (0.002)	0.016*** (0.005)
Dad's English Level					0.004 (0.005)	0.006 (0.004)	0.004 (0.006)	0.012* (0.007)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.06	0.11	0.14	0.16	0.14	0.17	0.18	0.17
R^2	0.06	0.16	0.19	0.28	0.27	0.29	0.29	0.34
Observations	250910	250910	250910	11619	11619	11619	7425	5705

Notes: Heteroskedasticity robust standard error estimates clustered at the parental countries of origin, language and state levels are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

tense has an effect that is about a third of the effect of having a college educated mother or above half the effect of having a college educated father.

Third, although the analysis accounts for parental origin fixed-effects and language level geographical characteristics as well as additional controls, the effect of periphrastic future tense may reflect the language-embodied effect of other (ancestral) cultural traits associated with an individual's language, which may be unrelated to long-term orientation and its potential reflection in the periphrastic future tense. Table 7 explores this possibility by additionally accounting for other language structures. Reassuringly, the effect of periphrastic future tense is unaffected by the inclusion of these additional language structures, which are mostly insignificant.

Finally, individual's educational choices can be affected by local socio-economic conditions. In particular, local labor market conditions and opportunities might be affected by ethnic networks, racial or ethnic discrimination, among others. The previous results addressed this issue partially by comparing observationally equivalent second-generation migrants within states. Table A.3 further establishes that the results are qualitatively similar if instead within-county level variation is exploited. Moreover,

Table 5: Periphrastic Future Tense and College Attendance of Second-Generation Migrants

	College Attendance					
	(1)	(2)	(3)	(4)	(5)	(6)
Periphrastic Future Tense	0.201*** (0.013)	0.207*** (0.007)	0.201*** (0.007)	0.046*** (0.011)		0.041*** (0.012)
Crop Return (pre-1500CE) (mean)					0.013*** (0.004)	0.007* (0.004)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	No	Yes	Yes	Yes	Yes
Parental Origin FE	No	No	No	Yes	Yes	Yes
Adjusted- R^2	0.05	0.07	0.08	0.13	0.13	0.13
R^2	0.05	0.11	0.17	0.45	0.45	0.45
Observations	165250	165250	165250	165250	165250	165250

Notes: This table establishes the negative significant effect of future tense on college attendance of second-generation migrants in the US. Heteroskedasticity robust standard error estimates clustered at the level of the included fixed-effects are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

labor market opportunities might be affected by speaking one of the two main languages in the US, namely English and Spanish. Additionally, the recent increase in (Spanish speaking) immigrants from Latin-America, many with lower levels of human capital, may bias the results. Reassuringly, Table 8 establishes that the results remain qualitatively unchanged if English or Spanish speakers are excluded from the analysis.

The previous results suggest that speaking a language with periphrastic future tense directly increases the probability of attending college. One potential interpretation of these results is that using the periphrastic future tense in itself affects behavior. On the other hand, it could be capturing within-country of origin variations in time preference. In particular, if parents come from the same country of origin, but differ in their culture and language, the effect of periphrastic future tense might just be capturing these cultural differences. Table A.4 shows the results of splitting the sample of second-generation migrants among those whose parents come from the same country and those whose parents come from different countries. The table establishes that periphrastic future tense has no effect in the sample of individuals whose parents come from the same country.⁸ On the contrary, the effect of periphrastic future tense remains qualitatively unchanged in the sample of migrants whose parents come from different countries. Although this could still capture some within-country of origin variation, it is less probable to do so. While the effect of periphrastic future tense may reflect the (transmitted) long-term orientation of the parent whose language is spoken at home, the analysis cannot refute the presence of a direct effect of this language structure on college attendance.

⁸This result is driven by the lack of variation in the existence of periphrastic future tense of the language spoken at home, and thus, the parental countries of origin fixed-effects absorb all the potential explanatory power of periphrastic future tense.

Table 6: Periphrastic Future Tense and College Education of Second Generation Migrants Accounting for Parental Education and English Levels

	College Attendance								
	Parental Education			Parental English			Both		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Periphrastic Future Tense	0.047*** (0.007)		0.043*** (0.007)	0.038*** (0.008)		0.038*** (0.009)	0.035*** (0.008)		0.034*** (0.009)
Crop Return (pre-1500CE)		0.013*** (0.003)	0.007*** (0.003)		0.005 (0.003)	0.000 (0.003)		0.006** (0.003)	0.003 (0.003)
Mother's College Attendance	0.130*** (0.003)	0.130*** (0.003)	0.130*** (0.003)				0.134*** (0.004)	0.134*** (0.004)	0.134*** (0.004)
Father's College Attendance	0.073*** (0.003)	0.073*** (0.003)	0.073*** (0.003)				0.146*** (0.004)	0.147*** (0.004)	0.146*** (0.004)
Mother's English Level				0.012*** (0.001)	0.012*** (0.001)	0.012*** (0.001)	0.013*** (0.001)	0.014*** (0.001)	0.013*** (0.001)
Father's English Level				-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	0.006*** (0.002)	0.006*** (0.002)	0.006*** (0.002)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE for Both Parents	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.14	0.14	0.14	0.14	0.14	0.14	0.18	0.18	0.18
R^2	0.23	0.23	0.23	0.23	0.23	0.23	0.26	0.26	0.26
Observations	165250	165250	165250	98623	98623	98623	98623	98623	98623

Notes: This table establishes the robustness of the positive effect of periphrastic future tense on college attendance to the inclusion of parental educational and English levels. Heteroskedasticity robust standard error estimates clustered at the level of the included fixed-effects are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

2.3 Plow Suitability, Sex-Based Grammatical Gender, and Education of Second-Generation Female Migrants

This section explores the effect of languages with sex-based grammatical gender on human capital accumulation of its speakers. In view of the proposed hypothesis that in a society characterized by distinct gender roles and consequently by the existence of gender bias, sex-based grammatical gender systems could have fortified the existing social structure and cultural norms, the analysis explores whether languages with sex-based grammatical gender have an adverse effect on the human capital accumulation of its female speakers. In particular, following the identification strategy exploited in the previous section, the analysis focuses on the effect of sex-based grammatical gender on college attendance of second-generation female migrants into the US.

In line with the proposed hypothesis, Table 9 establishes the negative effect of speaking a language with sex-based grammatical gender on college attendance of female speakers. In particular, accounting geographical characteristics of the homeland of the language, column (1) shows that women who speak a language with sex-based grammatical gender have 23 percentage points lower probability of

Table 7: Periphrastic Future Tense and College Education of Second Generation Migrants Accounting for other Linguistic Structures

	College Attendance						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Periphrastic Future Tense	0.043*** (0.007)	0.049*** (0.008)	0.041*** (0.008)	0.041*** (0.014)	0.045*** (0.008)	0.047*** (0.008)	0.048*** (0.008)
Crop Return (pre-1500CE)	0.007*** (0.003)	0.012*** (0.003)	0.011*** (0.003)	-0.004 (0.004)	0.010** (0.004)	0.015*** (0.004)	0.014*** (0.003)
Mother's College Attendance	0.130*** (0.003)	0.131*** (0.003)	0.131*** (0.003)	0.133*** (0.003)	0.132*** (0.003)	0.132*** (0.003)	0.132*** (0.003)
Father's College Attendance	0.073*** (0.003)	0.075*** (0.003)	0.075*** (0.003)	0.076*** (0.003)	0.076*** (0.003)	0.075*** (0.003)	0.075*** (0.003)
Past Tense		0.015 (0.014)					
Perfect Tense			-0.011 (0.007)				
Existence of Gender System				-0.030* (0.018)			
Evidentiality					0.018** (0.008)		
Consonant Inventories						0.001 (0.007)	
Consonant-Vowel Ratio							0.001 (0.004)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE for Both Parents	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.14	0.14	0.14	0.14	0.14	0.14	0.14
R^2	0.23	0.23	0.23	0.23	0.23	0.23	0.23
Observations	165250	158239	158239	153996	155905	157002	157002

Notes: This table establishes the robustness of the positive effect of periphrastic future tense on college attendance to accounting for other language structures. The analysis accounts for parental ancestry fixed-effect, as well as for parental college attendance. Heteroskedasticity robust standard error estimates clustered at the level of the included fixed-effects are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

attending college in comparison to women with similar observable characteristics who live in the same county and who were interviewed the same year. Nevertheless, this effect may capture the persistence of characteristics of the parental countries of origin of these women independently of grammatical gender. Thus, column (2) accounts for parental origins fixed-effects, and therefore isolates the effect of gender bias that is language-embodied from the persistent cultural effects of gender bias via non-linguistic channels. The results suggest that sex-based grammatical gender per se has an adverse effect on the probability of attending college, lowering this probability by 2.4 percentage points. Moreover, accounting for the geographical origins of sex-based grammatical gender, Column (3) suggests that the adverse effect of sex-based grammatical gender per se on the probability of attending remains significant. Finally, as established in columns (4) and (5), the adverse effect of speaking a language

Table 8: Periphrastic Future Tense and College Education of Second Generation Migrants Accounting for Local Labor Market Conditions

	College Attendance					
	No English			No Spanish		
	(1)	(2)	(3)	(4)	(5)	(6)
Periphrastic Future Tense	0.021** (0.009)		0.022** (0.009)	0.029*** (0.006)		0.027*** (0.006)
Crop Return (pre-1500CE) (mean)		0.001 (0.002)	-0.001 (0.002)		0.005*** (0.002)	0.002 (0.002)
Mother's College Attendance	0.114*** (0.007)	0.114*** (0.007)	0.114*** (0.007)	0.124*** (0.007)	0.123*** (0.007)	0.124*** (0.007)
Father's College Attendance	0.135*** (0.007)	0.135*** (0.007)	0.135*** (0.007)	0.131*** (0.007)	0.130*** (0.007)	0.131*** (0.007)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE for Both Parents	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.19	0.19	0.19	0.19	0.19	0.19
R^2	0.34	0.34	0.34	0.34	0.34	0.34
Observations	52537	52537	52537	55176	55176	55176

Notes: This table establishes the positive significant effect of periphrastic future tense on college attendance excluding English and Spanish speakers. The analysis accounts for parental ancestry fixed-effect, as well as for parental college attendance. Heteroskedasticity robust standard error estimates three-way clustered by state and country of origin of both parents are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

with sex-based grammatical gender on female college attendance is robust the the confounding effect of parental education. Thus, the analysis in Table 9 suggests that speaking a language with sex-based grammatical gender has an adverse effect on female college attendance.

3 Conclusion

This research explores some of the most fundamental and intriguing mysteries about the origins of the coevolution of linguistic and cultural traits and their impact on the development process: Has the coevolution of linguistic and cultural traits contributed to the persistence of cultural characteristics and their lasting effect on economic prosperity? Have language structures merely reflected existing cultural traits or have they influenced human behavior and values and contributed directly to the development process? What are the geographical roots of the coevolution of linguistic and cultural traits? Are the geographical characteristics that triggered the coevolution of culture and language critical for the understanding of the contribution of cultural and linguistic characteristics for the wealth of nations?

The study advances the hypothesis and establishes empirically that variations in pre-industrial geographical characteristics that were conducive to higher returns to agricultural investment, gender gaps in agricultural productivity, and hierarchical societies, are at the root of existing cross-language

Table 9: Gender and College Attendance of Female Second Generation Migrants

	Female College Attendance				
	(1)	(2)	(3)	(4)	(5)
Existence of Gender System	-0.228*** (0.018)	-0.024*** (0.009)	-0.017* (0.009)	-0.049*** (0.007)	-0.046*** (0.008)
Average Caloric Yield (Plow Negative Crops, pre-1500)			-0.013* (0.005)		-0.007 (0.004)
Average Caloric Yield (All Crops, pre-1500)			0.007 (0.003)		0.000 (0.003)
Mother's College Attendance				0.121*** (0.007)	0.121*** (0.007)
Father's College Attendance				0.137*** (0.007)	0.137*** (0.007)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes
Origin FE for Both Parents	No	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes
Marital Status FE	Yes	Yes	Yes	Yes	Yes
Parental Origin FE	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.08	0.13	0.13	0.16	0.16
R^2	0.24	0.28	0.28	0.31	0.31
Observations	39433	39433	39433	39433	39433

Notes: This table establishes the negative significant effect of sex-based grammatical gender on female college attendance. Heteroskedasticity robust standard error estimates three-way clustered by state and country of origin of both parents are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

variations in the structure of the future tense, and the presence of grammatical gender and politeness distinctions. Moreover, the research suggests that while language structures have largely reflected past human experience and in particular ancestral cultural traits in society, they have independently affected human behavior and economic outcomes.

The empirical methodology that is advanced in the course of this research augments the epidemiological approach and permits the analysis to isolate the effect of cultural traits that are language-embodied on human behavior from the persistent effects of culture via non-linguistic channels. In particular, it suggests that variations in the languages spoken by second-generation migrants originated from the same ancestral regions can be exploited to account for country of origin fixed-effects and thus to overcome the potential biases that could be generated by omitted ancestral characteristics.

References

- Acemoglu, D., Johnson, S. and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation, *The American Economic Review* **91**(5): 1369–1401.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S. and Wacziarg, R. (2003). Fractionalization, *Journal of Economic Growth* **8**(2): 155–194.
- Alesina, A. and Ferrara, E. L. (2005). Ethnic diversity and economic performance, *Journal of economic literature* **43**(3): 762–800.

- Ashraf, Q. and Galor, O. (2013). The out of africa hypothesis, human genetic diversity, and comparative economic development, *The American Economic Review* **103**(1): 1–46.
- Bybee, J. L. and Dahl, Ö. (1989). The creation of tense and aspect systems in the languages of the world, *Studies in language* **13**(1): 51–103.
- Bybee, J. L. and Pagliuca, W. (1987). The evolution of future meaning, *Papers from the 7th international conference on historical linguistics*, Amsterdam: John Benjamins, pp. 109–122.
- Bybee, J. L., Perkins, R. and Pagliuca, W. (1994). The evolution of grammar.
- Chen, M. K. (2013). The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets, *The American Economic Review* **103**(2): 690–731.
- Dahl, Ö. (1985). *Tense and aspect systems*, Basil Blackwell.
- Dahl, Ö. (2000). The grammar of future time reference in european languages, *Tense and Aspect in the Languages of Europe* p. 309.
- Dahl, Ö. and Velupillai, V. (2013). The future tense, in M. S. Dryer and M. Haspelmath (eds), *The World Atlas of Language Structures Online*, Max Planck Institute for Evolutionary Anthropology, Leipzig.
- Desmet, K., Ortuño-Ortín, I. and Wacziarg, R. (2012). The political economy of linguistic cleavages, *Journal of development Economics* **97**(2): 322–338.
- Duncan, B. and Trejo, S. J. (2011). Intermarriage and the intergenerational transmission of ethnic identity and human capital for mexican americans, *Journal of Labor Economics* **29**(2): 195.
- Duncan, B. and Trejo, S. J. (2016). The complexity of immigrant generations: Implications for assessing the socioeconomic integration of hispanics and asians, *NBER Working Paper Series* (w21982).
- Easterly, W. and Levine, R. (1997). Africa’s growth tragedy: policies and ethnic divisions, *The Quarterly Journal of Economics* pp. 1203–1250.
- Eckert, P. and McConnell-Ginet, S. (2003). *Language and gender*, Cambridge University Press, Cambridge.
- Fearon, J. D. (2003). Ethnic and cultural diversity by country, *Journal of Economic Growth* **8**(2): 195–222.
- Fernandez, R. and Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility, *American Economic Journal: Macroeconomics* **1**(1): 146–177.
- Figlio, D., Giuliano, P., Özek, U. and Sapienza, P. (2016). Long-term orientation and educational performance, *Technical report*, National Bureau of Economic Research.
- Gallup, J. L., Sachs, J. D. and Mellinger, A. D. (1999). Geography and economic development, *International regional science review* **22**(2): 179–232.
- Galor, O. and Özak, Ö. (2016). The agricultural origins of time preference, *American Economic Review* **106**(10): 3064–3103.
- Giuliano, P. (2007). Living arrangements in western europe: Does cultural origin matter?, *Journal of the European Economic Association* **5**(5): 927–952.
- Glaeser, E. L., La Porta, R., Lopez-de Silanes, F. and Shleifer, A. (2004). Do institutions cause growth?, *Journal of economic Growth* **9**(3): 271–303.
- Guiso, L., Sapienza, P. and Zingales, L. (2004). The role of social capital in financial development, *American Economic Review* **94**(3): 526–556.
- Guiso, L., Sapienza, P. and Zingales, L. (2006). Does culture affect economic outcomes?, *Journal of Economic Perspectives* **20**(2): 23–48.
- Harutyunyan, A. and Özak, Ö. (2016). Culture, diffusion, and economic development, *SMU Working Paper Series* .
- Lakoff, R. (1973). Language and woman’s place, *Language in society* **2**(1): 45–79.
- Roberts, S. G., Winters, J. and Chen, K. (2015). Future tense and economic decisions: controlling for cultural evolution, *PLoS one* **10**(7): e0132145.

- Ruggles, S., Genadek, K., Goeken, R., Grover, J. and Sobek, M. (2015). *Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]*, University of Minnesota, Minneapolis.
- Stahlberg, D., Braun, F., Irmen, L. and Sczesny, S. (2007). Representation of the sexes in language, *Social communication* pp. 163–187.
- Tabellini, G. (2010). Culture and institutions: economic development in the regions of europe, *Journal of the European Economic Association* **8**(4): 677–716.

Appendix (Not for publication)

A Language Structures & Contemporary Behavior

A.1 Identification Strategy

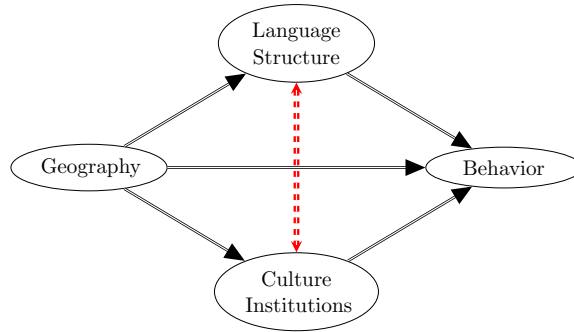


Figure A.1: Language and Contemporary Behavior

A.2 Crop Return, Periphrastic Future Tense, and Long-Term Oriented Behavior of Second-Generation Migrants

Table A.1: Pre-1500CE Crop Return, Periphrastic Future Tense, and College Education of Second and Higher Generation Migrants

	College Attendance				
	(1)	(2)	(3)	(4)	(5)
Periphrastic Future Tense	0.100*** (0.014)	0.132*** (0.005)	0.125*** (0.004)		0.111*** (0.004)
Crop Return (pre-1500CE)				0.034*** (0.001)	0.019*** (0.001)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes
Marital Status FE	No	No	Yes	Yes	Yes
Adjusted- R^2	0.02	0.07	0.08	0.08	0.08
R^2	0.02	0.07	0.09	0.09	0.09
Observations	12206839	12206839	12206839	12206839	12206839

Table A.2: Pre-1500CE Crop Return, Periphrastic Future Tense, and College Education
One-and-a-Half Generation Migrants (Who Arrived at Age ≤ 5)

	College Attendance					
	Language				Crop Return	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Periphrastic Future Tense	0.205*** (0.013)	0.208*** (0.007)	0.204*** (0.005)	0.056*** (0.007)		0.054*** (0.007)
Crop Return (pre-1500CE)					0.011*** (0.003)	0.004 (0.003)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	No	Yes	Yes	Yes	Yes
Parental Origin FE	No	No	No	Yes	Yes	Yes
Adjusted- R^2	0.06	0.09	0.10	0.15	0.15	0.15
R^2	0.06	0.11	0.17	0.48	0.48	0.48
Observations	422081	422081	422081	422081	422081	422081

Notes: Heteroskedasticity robust standard error estimates clustered at the level of the included fixed-effects are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table A.3: Periphrastic Future Tense and College Education of Second Generation Migrants Accounting for Parental Education and English Levels, and Local Socio-Economic Conditions (County Level)

	College Attendance								
	Parental Education			Parental English			Both		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Periphrastic Future Tense	0.032*** (0.005)		0.029*** (0.005)	0.030*** (0.006)		0.030*** (0.006)	0.029*** (0.005)		0.027*** (0.005)
Crop Return (pre-1500CE)		0.006*** (0.002)	0.003** (0.001)		0.003* (0.002)	0.000 (0.001)		0.005*** (0.001)	0.003** (0.001)
Mother's College Attendance	0.131*** (0.006)	0.131*** (0.006)	0.131*** (0.006)				0.133*** (0.006)	0.133*** (0.006)	0.133*** (0.006)
Father's College Attendance	0.141*** (0.006)	0.142*** (0.006)	0.141*** (0.006)				0.143*** (0.006)	0.143*** (0.006)	0.143*** (0.006)
Mother's English Level				0.013*** (0.002)	0.013*** (0.002)	0.013*** (0.002)	0.014*** (0.001)	0.014*** (0.001)	0.014*** (0.001)
Father's English Level				-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	0.006*** (0.001)	0.006*** (0.001)	0.006*** (0.001)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE for Both Parents	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.18	0.18	0.18	0.14	0.14	0.14	0.18	0.18	0.18
R^2	0.30	0.30	0.30	0.27	0.27	0.27	0.30	0.30	0.30
Observations	91613	91613	91613	91613	91613	91613	91613	91613	91613

Notes: Heteroskedasticity robust standard error estimates three-way clustered by state and country of origin of both parents are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table A.4: Pre-1500CE Crop Return, Periphrastic Future Tense, and College Education of Second Generation Migrants
Effect of Parents Origin

	College Attendance					
	Same			Different		
	(1)	(2)	(3)	(4)	(5)	(6)
Periphrastic Future Tense	0.009 (0.006)		0.010 (0.006)	0.028*** (0.008)		0.014* (0.008)
Crop Return (pre-1500CE)		-0.000 (0.001)	-0.001 (0.001)		0.018*** (0.003)	0.016*** (0.003)
Mother's College Attendance	0.112*** (0.005)	0.112*** (0.005)	0.112*** (0.005)	0.155*** (0.006)	0.155*** (0.006)	0.155*** (0.006)
Father's College Attendance	0.121*** (0.006)	0.121*** (0.006)	0.121*** (0.006)	0.163*** (0.007)	0.163*** (0.007)	0.163*** (0.007)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	Yes	Yes	Yes	Yes	Yes	Yes
Parental Origin FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.19	0.19	0.19	0.18	0.18	0.18
R^2	0.34	0.34	0.34	0.37	0.37	0.37
Observations	54252	54252	54252	42614	42614	42614

Notes: Heteroskedasticity robust standard error estimates three-way clustered by state and country of origin of both parents are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table A.5: Pre-1500CE Crop Return, Periphrastic Future Tense, and College Education of Second Generation Migrants

	Educational Level Higher than High School					
	(1)	(2)	(3)	(4)	(5)	(6)
Periphrastic Future Tense	0.218*** (0.013)	0.225*** (0.008)	0.221*** (0.007)	0.056*** (0.012)		0.053*** (0.012)
Crop Return (pre-1500CE)					0.008*** (0.003)	0.003 (0.003)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	No	Yes	Yes	Yes	Yes
Parental Origin FE	No	No	No	Yes	Yes	Yes
Adjusted- R^2	0.05	0.12	0.20	0.49	0.49	0.49
R^2	0.05	0.12	0.20	0.49	0.49	0.49
Observations	18845303	18845303	18845303	18845303	18845303	18845303

Notes: Heteroskedasticity robust standard error estimates clustered at the level of the included fixed-effects are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table A.6: Pre-1500CE Crop Return, Periphrastic Future Tense, and College Education of Second Generation Migrants
Accounting for Parental Education and English Levels

	Educational Level Higher than High School								
	Parental Education			Parental English			Both		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Periphrastic Future Tense	0.056*** (0.008)		0.052*** (0.009)	0.041*** (0.010)		0.042*** (0.011)	0.037*** (0.010)		0.038*** (0.011)
Crop Return (pre-1500CE)		0.010*** (0.002)	0.004* (0.002)		0.002 (0.003)	-0.002 (0.003)		0.003 (0.003)	-0.000 (0.003)
Mother's Education Level (HS+)	0.137*** (0.004)	0.137*** (0.004)	0.137*** (0.004)				0.138*** (0.005)	0.138*** (0.005)	0.138*** (0.005)
Father's Education Level (HS+)	0.069*** (0.004)	0.069*** (0.004)	0.069*** (0.004)				0.147*** (0.005)	0.147*** (0.005)	0.147*** (0.005)
Mother's English Level				0.012*** (0.002)	0.012*** (0.002)	0.012*** (0.002)	0.013*** (0.002)	0.013*** (0.002)	0.013*** (0.002)
Father's English Level				-0.002 (0.002)	-0.002 (0.002)	-0.002 (0.002)	0.003* (0.002)	0.003* (0.002)	0.003* (0.002)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Origin FE for Both Parents	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parental Origin FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	0.25	0.25	0.25	0.26	0.26	0.26	0.29	0.29	0.29
R^2	0.25	0.25	0.25	0.26	0.26	0.26	0.29	0.29	0.29
Observations	18845303	18845303	18845303	11187136	11187136	11187136	11187136	11187136	11187136

Notes: Heteroskedasticity robust standard error estimates clustered at the level of the included fixed-effects are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

Table A.7: Pre-1500CE Crop Return, Periphrastic Future Tense, and College Education
One-and-a-Half Generation Migrants (Who Arrived at Age ≤ 5)

	Educational Level Higher than High School					
	(1)	(2)	(3)	(4)	(5)	(6)
Periphrastic Future Tense	0.220*** (0.011)	0.223*** (0.006)	0.219*** (0.006)	0.067*** (0.007)		0.067*** (0.007)
Crop Return (pre-1500CE)					0.007*** (0.002)	-0.000 (0.002)
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	No	Yes	Yes	Yes	Yes	Yes
Gender FE	No	Yes	Yes	Yes	Yes	Yes
Marital Status FE	No	No	Yes	Yes	Yes	Yes
Parental Origin FE	No	No	No	Yes	Yes	Yes
Adjusted- R^2	0.06	0.12	0.20	0.55	0.55	0.55
R^2	0.06	0.12	0.20	0.55	0.55	0.55
Observations	42457006	42457006	42457006	42457006	42457006	42457006

Notes: Heteroskedasticity robust standard error estimates clustered at the level of the included fixed-effects are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

A.3 Comparing Migrants Samples

Table A.8: Means across Generational Samples (Census vs. CPS)

	Means							
	1.5 Generation		2nd Generation				2+ Generations	
	Census	CPS	Census	CPS (living with Parents)	CPS (not liv- ing with Parents)	CPS (All)	Census	CPS (3+ Genera- tion)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Education Level (HS+)	0.596*** (0.001)	0.648*** (0.001)	0.552*** (0.001)	0.600*** (0.002)	0.568*** (0.000)	0.571*** (0.000)	0.535*** (0.000)	0.572*** (0.000)
Age	43.742*** (0.022)	38.625*** (0.024)	33.913*** (0.022)	34.092*** (0.032)	55.963*** (0.017)	54.376*** (0.017)	51.685*** (0.004)	50.133*** (0.004)
Gender	1.518*** (0.001)	1.518*** (0.001)	1.457*** (0.001)	1.462*** (0.002)	1.537*** (0.000)	1.531*** (0.000)	1.526*** (0.000)	1.527*** (0.000)
Marital Status	2.702*** (0.003)	2.737*** (0.005)	4.933*** (0.004)	5.099*** (0.005)	2.597*** (0.002)	2.779*** (0.002)	2.524*** (0.000)	2.489*** (0.001)
Observations	429372	174094	181099	94331	1205633	1299964	20596324	14180541

Notes: Standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

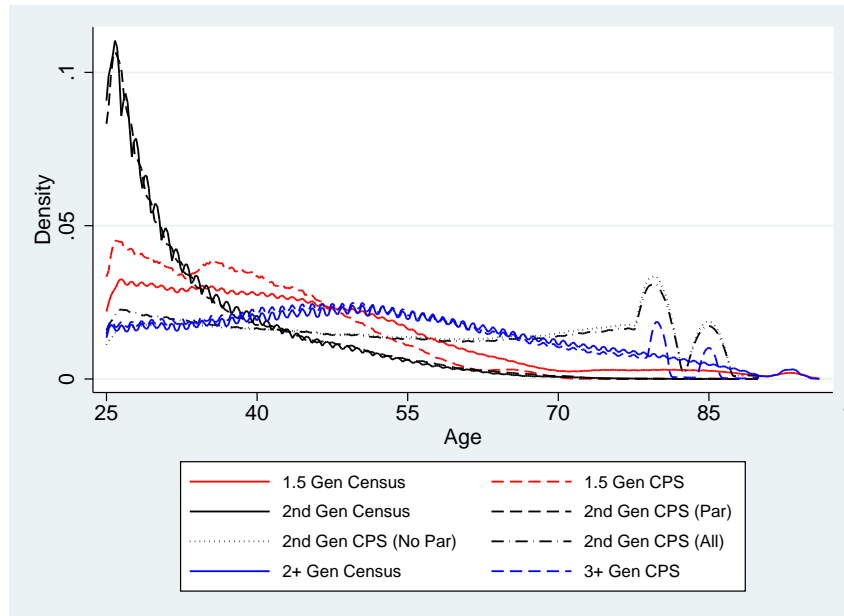


Figure A.2: Age Density in the Census and CPS

Table A.9: Means across Generational Samples (Census vs. CPS)

	Means							
	1.5 Generation		2nd Generation				2+ Generations	
	Census	CPS	Census	CPS (living with Parents)	CPS (not liv- ing with Parents)	CPS (All)	Census	CPS (3+ Genera- tion)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Education Level (HS+)	0.590*** (0.000)	0.638*** (0.000)	0.545*** (0.000)	0.603*** (0.000)	0.581*** (0.000)	0.583*** (0.000)	0.542*** (0.000)	0.568*** (0.000)
Age	42.489*** (0.002)	38.163*** (0.000)	33.347*** (0.002)	33.505*** (0.001)	54.517*** (0.000)	52.666*** (0.000)	50.258*** (0.000)	49.930*** (0.000)
Gender	1.510*** (0.000)	1.503*** (0.000)	1.445*** (0.000)	1.437*** (0.000)	1.530*** (0.000)	1.521*** (0.000)	1.522*** (0.000)	1.524*** (0.000)
Marital Status	2.855*** (0.000)	2.810*** (0.000)	5.000*** (0.000)	5.143*** (0.000)	2.624*** (0.000)	2.846*** (0.000)	2.679*** (0.000)	2.547*** (0.000)
Observations	43181154	403711034	20841131	245898566	2544423483	2790322049	1831557413	28887227869

Notes: Standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

A.4 Future Tense and Education in the WVS

Table A.10: Pre1500 Crop Return, Periphrastic Future Tense, and Education – World Values Survey

	Education Level								
	Basic Controls			Income			Religion		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Crop Return (pre-1500CE)	0.18*** (0.04)		0.23*** (0.04)	0.18*** (0.04)		0.23*** (0.04)	0.25*** (0.04)		0.28*** (0.04)
Periphrastic Future Tense		0.43*** (0.05)	0.47*** (0.05)		0.40*** (0.05)	0.44*** (0.05)		0.25*** (0.05)	0.30*** (0.05)
Income FE	No	No	No	No	No	No	Yes	Yes	Yes
Religion FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Main Geographical Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted- R^2	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.09	0.09	0.09
Observations	108213	108213	108213	108213	108213	108213	108213	108213	108213

B Variable Definitions, Sources and Summary Statistics

Table B.11: Summary Statistics of the Existence of Periphrastic Future Tense by Region

Region	Observations	Mean	Std. Dev.
Sub-Saharan Africa	66	0.53	0.503
Middle East and North Africa	8	0.5	0.53
Europe and Central Asia	56	0.48	0.50
South Asia	21	0.19	0.40
East Asia and Pacific	71	0.55	0.50
North America	22	0.41	0.50
Latin America	31	0.55	0.50
Total	275	0.49	0.50

Table B.12: Summary Statistics of the Existence of Sex-Based Grammatical Gender Systems by Region

Region	Observations	Mean	Std. Dev.
Sub-Saharan Africa	27	0.63	0.49
Middle East and North Africa	7	0.71	0.49
Europe and Central Asia	40	0.48	0.51
South Asia	16	0.63	0.50
East Asia and Pacific	70	0.27	0.45
North America	25	0.08	0.28
Latin America	32	0.28	0.46
Total	227	0.37	0.48

Table B.13: Summary Statistics of the Existence of Politeness Distinctions by Region

Region	Observations	Mean	Std. Dev.
Sub-Saharan Africa	36	0.14	0.35
Middle East and North Africa	4	0.25	0.50
Europe and Central Asia	34	0.71	0.46
South Asia	19	0.63	0.50
East Asia and Pacific	59	0.32	0.47
North America	18	0.00	0.00
Latin America	28	0.18	0.39
Total	207	0.34	0.48

Table B.14: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Absolute Latitude	0.096	1.025	-1.302	2.613	275
Elevation	0.027	1.026	-0.92	4.827	275
Ruggedness	-0.014	0.979	-0.877	6.162	275
Coast Length	0.024	1.154	-0.302	11.692	275
Precipitation	-0.078	0.928	-1.3	4.4	275
Precipitation (std)	-0.02	0.911	-0.667	8.314	275
Precipitation Volatility	-0.064	0.926	-1.531	4.665	275
Precipitation Spatial Correlation	0.064	0.939	-2.133	0.810	275
Temperature (Daily Mean)	-0.054	0.977	-2.996	1.176	275
Temperature (Daily Mean) (std)	-0.017	0.929	-0.877	4.876	275
Temperature Volatility	0.079	0.991	-1.641	3.504	275
Temperature Spatial Correlation	0.068	0.939	-2.161	0.683	275

Table B.15: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Intensity of Agriculture	8.890	3.061	2	12	264