Economics Letters 168 (2018) 46–48

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Gendered language and the educational gender gap

Lewis Davis*, Megan Reynolds[1]

Economics Department, Union College, 807 Union Street, Schenectady, NY 12308, United States

HIGHLIGHTS

- Speaking a gendered language is associated with a 0.75 year increase in the educational gender gap.
- It is also associated with a 7.6% point rise in the gender gap in secondary school completion.

ARTICLE INFO

Article history:
Received 3 January 2018
Received in revised form 2 April 2018
Accepted 4 April 2018
Available online 11 April 2018

JEL classification:
I24
J16
Z1

Keywords:
Culture
Language
Education
Gender inequality

ABSTRACT

Languages differ in the degree to which they employ gender distinctions for nouns and pronouns. Speaking a gendered language may highlight gender roles. We find that speaking a gendered language is associated with a greater gender gap in educational attainment.

1. Introduction

Languages differ in the degree to which gender plays a role in their grammatical structure. According to linguistic relativity theory (Whorf, 1956), speaking a more heavily gendered language may highlight gender distinctions in the mind of the speaker, leading to more pronounced gender roles and greater disparities in social outcomes across genders.[3] Countries in which the dominant language is more gendered have lower rates of female participation in labor and credit markets (Gay et al., 2013; Mavisakalyan, 2015) and are more likely to have political gender quotas (Santacreu-Vasut et al., 2013). At the individual level, speaking a gendered language is associated with lower rates of female labor force participation (Gay et al., 2013, 2015), early female marriage (Gay et al., 2013), and a gendered division of labor in household tasks (Hicks et al., 2014).

We contribute to this literature by examining the relationship between gendered language and the educational gender gap. Using data from the World Values Survey, we find that a unit increase in our primary measure of gendered language, a dummy variable for whether a language has two sex-based noun classes, is associated with a 0.75 year decrease in educational attainment and a 7.6 percentage point decline in the secondary school completion rate among women in a given language-country group relative to similar men.

2. Data

Data on gendered language are from World Atlas of Language Structures and are summarized by three characteristics of a language's grammar that reflect the gender intensity of a language's nouns and pronouns. Our primary measure of gendered language is *sex_based_nouns*, a dummy variable for whether a language has two sex-based noun classes (Corbett, 2013a, b). This variable takes a value of one for Spanish, for example, which has masculine and feminine noun classes, zero for English, which has a single noun

* Corresponding author.
E-mail address: davisl@union.edu (L. Davis).

[1] Present address: Goldman Sachs, 200 West St, New York, NY 10282.

[3] Research in linguistics has linked gendered language to cognitive processes (Boroditsky et al., 2003) links and information processing (Vitevitch et al., 2013).

https://doi.org/10.1016/j.econlet.2018.04.006
0165-1765/© 2018 Published by Elsevier B.V.

class, and zero for German, which has three sex-based noun classes, masculine, feminine and neuter.[4]

Our second measure of the gender intensity of nouns reflects the system for assigning nouns to different classes, which may be either semantic or both semantic and formal (Corbett, 2013c). Semantic systems of gender assignment are based on a noun's meaning, e.g. nouns denoting males are masculine. Formal systems of gender assignment reflect formal linguistic elements like noun phonology or morphology. Formal systems of gender assignment tend to be more extensive, relegating fewer nouns to a default gender, and may reflect an "older" gender system. The dummy variable *formal_system* equals one if a languages employs both semantic and formal systems of gender assignment and zero otherwise. For example, this variable takes a value of one for Russian, for which a noun's gender is indicated by the pattern of inflection for different cases (nominative, accusative, etc.).

The third measure of gendered language reflects the gender intensity of a language's pronouns (Siewierska, 2013). The variable *gen_pron* equals one if a language has gender distinctions in the third person singular personal pronoun, equals two if the first or second person singular personal pronoun is also gendered, and zero otherwise (Mavisakalyan, 2015). These three measures of gendered language are positively correlated, with correlation coefficients of 0.69 and higher, suggesting that these three aspects of gendered language are closely related. Linguistic variables are matched to individual respondents in the World Values Survey (WVS) using the language spoken at home, which is available for the first three waves of the survey.

Measures of the educational gender gap are constructed using data from the WVS. We consider two measures of educational attainment, years of education, which reflect the opportunity cost of an individual's time, and a dummy variable for secondary school completion, which signifies the acquisition of a valuable labor market credential (Schneider, 2010). We ascribe three years of education to each of four categories of schooling (some primary, primary, some secondary and secondary) and two years to the remaining two categories (some tertiary and tertiary). The dependent variable, *edyearsgap*, is the difference between years of education for a respondent and the average years of education of men in her country-language group. The dummy variable *second* takes a value of one if an individual has completed vocational or university-preparatory secondary school. The variable *secondgap* is the difference between *second* for the respondent and the secondary school completion rate for men in her country-language group.

The sample is restricted to women belonging to country-language pairs with at least 100 observations. The resulting dataset consists of over 76,000 women speaking 34 languages, residing in 70 countries, and belonging to 99 distinct country-language groups.[5] Table 1 provides summary statistics for selected variables.

3. Results

Our empirical model is:

$$edgap_{icl} = \beta genlang_l + \theta X_i + \alpha_c + \gamma_t + \varepsilon_{icl}r \qquad (1)$$

where *edgap* is a measure of the educational gender gap, *genlang* is a measure of gendered language, and i, c and l index individuals, countries and languages. Country fixed effects, α_c, control for country-level omitted variables that might affect a woman's educational decisions, such as characteristics of educational and labor market institutions. Wave fixed effects, γ_t, control for global shocks to educational gender inequality. The vector of individual level characteristics, X_i, consists of age and age-squared. We cluster standard errors by country-language groups.

Female educational attainment is influenced by assortative matching and intergenerational transmission effects (Becker, 1985; Heineck and Riphahn, 2016). While the absence of data on spousal and parental educational attainment limits modeling these relationships, the language a woman speaks at home is likely to be correlated with the languages spoken by her parents and spouse. Because of this, the coefficient on the gendered language variable is likely to reflect a variety of effects, including those operating through parental and spousal influences.

Mavisakalyan (2015) and Davis and Williamson (2016) address the endogeneity of language structures using instruments derived from the structures of related languages. However, Galor et al. (2018) argue that language structures evolve to complement existing cultural norms, a claim that raises questions about the validity of such instruments, as they are likely to be correlated with unobserved dimensions of culture related to gender roles. Because of this we do not attempt to address the endogeneity of gendered language.

Panel A of Table 2 shows results for the gender gap in years of education. In our baseline specification, we focus on whether a language has two sex-based noun classes, as reflected by *sex_based_nouns*. This measure is negative and significant at the 1% level. The coefficient estimate indicates that speaking a language with two sex-based noun classes is associated with a 0.75-year increase in the educational gender gap, which equals 21% of the standard deviation of the gender gap in years of education. In the next three specifications we consider the other measures of gendered language. While both *formal_system* and *gen_pron* are significant when entered separately, only *sex_based_nouns* is significant when all three measures are entered simultaneously.

Languages are not evenly distributed across or within countries, raising the question of whether our results are driven in part by languages that are prominent either globally or within particular countries. We address this issue by considering two subsamples. First, we exclude from the sample the speakers of four global languages, Arabic, English, Russian and Spanish, which together comprise 64.1% of the sample. Second, we restrict the sample to speakers of minority languages, which excludes 69.5% of the sample. In both regressions, we focus on our primary measure gendered language, *sex_based_nouns*. As seen in columns 5 and 6, these sample restrictions modestly increase the estimated effect of gendered language on gender gap in years of education.

In Table 2(B), we repeat these exercises using *secondgap* as the dependent variable. Our primary measure of gendered language, *sex_based_nouns*, is negative and statistically significant in each specification. The estimate in our baseline specification indicates that speaking a language with two sex-based noun classes is associated with a 7.6 percentage point decrease in female secondary school completion rates, relative to men, which equals 16.6% of a standard deviation of this variable. The significance of *sex_based_nouns* is robust to the inclusion of other measures of gendered language and to restricting the sample non-global and minority languages. These alternative specifications also have little impact on the magnitude of the estimated coefficient.

[4] We also considered whether speaking a language with three sex-based noun classes affects the educational gender gap. Including a dummy variable for such languages in our baseline specification does not significantly affect the coefficient on *sex_based_nouns*. Moreover, the coefficient on the dummy for having three-sex-based noun classes is not significant.

[5] The languages are Amharic, Arabic, Armenian, Berber, Cantonese, English, Ewe, Finnish, French, Georgian, German, Hausa, Hindi, Hungarian, Igbo, Indonesian, Kannada, Kirghiz, Latvian, Mandarin, Marathi, Persian, Russian, Shona, Spanish, Swahili, Tagalog, Tamil, Thai, Turkish, Ukrainian, Uzbek, Yoruba, and Zulu.

Table 1
Summary statistics.

Variable	Obs.	Mean	Std. Dev.	Min	Max
sex_based_nouns	76,605	0.4323739	0.4954088	0	1
formal_system	76,605	0.62056	0.4852508	0	1
gen_pron	71,481	1.228396	0.747325	0	2
edyearsgap	76,605	−0.3600767	3.552062	−11.05531	8.306122
secondgap	76,605	−0.0294401	0.4538881	−0.95	0.7465438

Table 2
Gendered language and educational gender inequality.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample	Full sample	Full sample	Full sample	Excluding global languages	Minority languages
Panel A	edyearsgap	edyearsgap	edyearsgap	edyearsgap	edyearsgap	edyearsgap
sex_based_nouns	-0.745^{***}			-0.579^{***}	-0.895^{***}	-0.935^{***}
	(−4.628)			(−3.586)	(−3.422)	(−3.832)
formal_system		-0.453^{***}		0.0505		
		(−2.699)		(0.335)		
gen_pron			-0.461^{**}	−0.277		
			(−2.162)	(−1.540)		
age, age_sqr	yes	yes	yes	yes	yes	yes
countries	yes	yes	yes	yes	yes	yes
waves	yes	yes	yes	yes	yes	yes
Panel B	secondgap	secondgap	secondgap	secondgap	secondgap	secondgap
sex_based_nouns	-0.0756^{***}			-0.0733^{***}	-0.0737^{**}	-0.0836^{***}
	(−3.942)			(−3.661)	(−2.412)	(−2.955)
formal_system		-0.0423^{**}		0.0164		
		(−2.131)		(0.942)		
gen_pron			−0.0416	−0.0228		
			(−1.455)	(−0.886)		
Observations	76,605	76,605	71,481	71,481	27,531	23,381
R-squared	0.115	0.115	0.111	0.111	0.139	0.138
Number of clusters	99	99	89	89	53	60

Notes. Sample restricted to women and to languages with more than 100 speakers in a given country. All regressions control for a quadratic relationship in age and for country and wave fixed effects. In column 5, the sample excludes Arabic, English, Russian and Spanish speakers. In column 6, the sample is restricted to minority languages. Standard errors clustered by language. Robust t-statistics are in parentheses. Asterisks indicate statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4. Conclusion

This paper provides evidence that speaking a language with two sex-based noun classes is strongly associated with an increase in the gender gap in educational attainment, as measured by years of education and secondary school completion rates. Our results are consistent with the idea that gender distinctions in language increase the salience of traditional gender roles in the mind of the speaker and contribute to unequal social outcomes across genders. As is well known, the gender gap in education has been shrinking globally over the past several decades, in part due to the empowering effect of economic development (Duflo, 2012). Our analysis poses a cautionary counter-point to this trend, suggesting that some portion of educational gender inequality is linked to highly stable linguistic structures and, thus, may persist even as countries develop. Put differently, if language plays a role in the persistence of culturally defined gender norms, then attempts to reform the gender structure of language may be warranted (McCoubrey, 2017).

References

Becker, Gary, 1985. Human capital, effort, and the sexual division of labor. J. Labor Econ. 3 (1), S33–S58. Part 2.

Boroditsky, Lera, Schmidt, Lauren A., Phillips, Webb, 2003. Sex, syntax, and semantics. In: Getner, Dedre, Goldin-Meadow, Susan (Eds.), Language in Mind: Advances in the Study of Language and Thought. MIT Press, pp. 61–79.

Corbett, Greville G., 2013a. Number of genders. In: Dryer, Matthew S., Haspelmath, Martin (Eds.), The World Atlas of Language Structures Online. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Corbett, Greville G., 2013b. Sex-based and non-sex-based gender systems. In: Dryer, Matthew S., Haspelmath, Martin (Eds.), The World Atlas of Language Structures Online. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Corbett, Greville G., 2013c. Systems of gender assignment. In: Dryer, Matthew S., Haspelmath, Martin (Eds.), The World Atlas of Language Structures Online. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Davis, Lewis, Williamson, Claudia, 2016. Culture and the regulation of entry. J. Comparative Econom. 44 (4), 1055–1083.

Duflo, Esther, 2012. Women empowerment and economic development. J. Econ. Lit. 50 (4), 1051–1079.

Galor, Oded, Ozak, Omer, Sarid, Assaf, 2018. Geographical Origins of Language Structures. SSRN #3097220.

Gay, Victor, Hicks, Daniel L., Santacreu-Vaust, Estefania, Amir, Shohoam, 2015. Decomposing culture: Can gendered language influence women's economic engagement? 1 (1).

Gay, V., Santacreu-Vasut, E., Shoham, A., 2013. The grammatical origins of gender roles. In: Work. pap. Berkeley Econ. Hist. Lab. Pap. Ser.

Heineck, G., Riphahn, R., 2016. Intergenerational transmission of educational attainment in Germany — the last five decades. Jahrb. Natl. Stat. 229 (1), 36–60.

Hicks, Daniel L., Santacreu-Vasut, Esterfania, Amnir, Shoham, 2014. Does mother tongue make for women's work? J. Econ. Behav. Organ. 110 (2), 19–44.

Mavisakalyan, Astghik, 2015. Gender in language and gender in employment. Oxf. Dev. Stud. 43 (4), 403–424.

McCoubrey, Carmel, 2017. Toppling the grammar patriarchy. The New York Times, Nov. 16.

Santacreu-Vasut, E., Shoham, A., Gay, V., 2013. Do female/male distinctions in language matter? Evidence from gender political quotas. Appl. Econ. Lett. 20 (5), 495–498.

Schneider, Silke, 2010. Nominal comparability is not enough: (In-)equivalence of construct validity of cross-national measures of educational attainment in the European Social Survey. Res. Soc. Stratif. Mobil. 28 (3), 343–357.

Siewierska, Anna, 2013. Gender distinctions in independent personal pronouns. In: Dryer, Matthew S., Haspelmath, Martin (Eds.), The World Atlas of Language Structures Online. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Vitevitch, M.S., Sereno, J., Jongman, A., Goldstein, R., 2013. Speaker sex influences processing of grammatical gender. PLoS ONE 8 (11), e79701.

Whorf, Benjamin L., 1956. Language, Thought and Reality. MIT Press, Cambridge, MA.