I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14539

Economics of Minority Groups: Labour Market Returns and Transmission of Indigenous Languages

Diego de la Fuente Stevens
Panu Pelkonen

JULY 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14539

Economics of Minority Groups: Labour Market Returns and Transmission of Indigenous Languages

Diego de la Fuente Stevens
University of Sussex

Panu Pelkonen
University of Sussex and IZA

JULY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

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

www.iza.org

ABSTRACT

Economics of Minority Groups: Labour Market Returns and Transmission of Indigenous Languages

This study demonstrates a series of links between minority language skills, their economic return and their transmission across generations. Using a detailed matching procedure and different data sources, we estimate the likelihood of being employed for bilingual versus monolingual men for a large number of Mexican indigenous groups. We find that for indigenous groups, retaining the minority language along with Spanish increases employment opportunities. Furthermore, we show that the languages that are associated with larger labour market benefits are more likely to be passed on from parents to children, controlling for other factors. Overall, this study shows that the continuity of minority languages across generations is linked to concrete economic benefits, labour market specialisation, and insurance value, along with the usual social factors within the family and the community.

JEL Classification: J4, J15, J31, O54, Z1, Z13

Keywords: intergenerational transmission, language skills, bilingualism, return to skills, minority languages, indigenous group

Corresponding author:
Panu Pelkonen
University of Sussex
Brighton, BN1 9RF
United Kingdom

E-mail: p.o.pelkonen@sussex.ac.uk

1. Introduction

Many developing countries are characterised by a dominant official language[1] and one or many minority languages that have varying numbers of speakers and differing levels of official status. In the developed world, smaller languages are often beneficiaries of substantial legal protection and educational infrastructure that shield them from an erosion of speakers. This is typically not the case in developing and emerging economies where, in the absence of supportive state institutions, languages either sustain themselves in various networks, or don't and face extinction.

Current trends in linguistic diversity around the world suggest that minority languages are disappearing fast, and that 90% of the world's languages are expected to disappear in the next 100 years (Nettle and Romaine, 2002). For Mexico, the most important reasons for the loss of native languages in the past has been a 'forced language shift', an official policy favouring Spanish; but, in current times, increasingly also a 'voluntary language shift', an abandonment of the language even in the absence of its suppression. This raises questions as to why some languages survive while others fade to exist.

In families where at least one parent speaks an indigenous language, parents weigh the options on what languages to teach to children. In a typical case, parents face a choice of raising bilingual children (who speak both the minority and the majority language), or monolingual children (the majority language only). In the common case where the school system supports only the majority language, parents must make an active effort to maintain the minority language at home.

In this study, we examine the intergenerational transmission of language and the economic rationale of language choice by breaking the family decision to raise bilingual children to different effects. Namely, the strength of social networks and the expected economic return to knowing a minority language. The main social networks are the family and the local area. For the expected economic benefits, we estimate the effect of minority language bilingualism on employment likelihood and earnings.

The study focuses on Mexico, which is one of a handful of large countries which has a rich

[1] Such as English, Spanish or French.

tapestry of minority languages (66 currently spoken). In the setting, languages remain geographically clustered, which makes it unique to look at language variation and compare employment returns, transmission rates, their varying characteristics and conditions. As the Mexican censuses and income surveys distinguish individuals by both ethnicity and language, these may also be unique in allowing for a good documentation of the transmission of languages and their employment effects.

Firstly, we show that indigenous men who are bilingual have, on average, a 2-4 percentage points higher likelihood of being employed, and receive 4.9 percentage higher earnings, as opposed to observationally equivalent indigenous men who can only speak Spanish. These estimates are based on a combination of matching and least squares regression using censuses for employment returns, and the income and expenditure survey (ENIGH) for the earnings estimates. Importantly, the results between OLS and matching differ, suggesting that OLS fails to adequately account for the omitted variable bias due to socioeconomic status.

The large number of observations inherent to the censuses allows us to recover language specific employment returns and intergenerational transmission rates. With this, we show that the estimated employment return varies by language but remains positive for all language groups for which the estimate is statistically different to zero. We present evidence showing that this employment effect derives partly from a larger likelihood of the speakers of indigenous languages to work in agriculture. Results also point that language returns are increasing with the intensity of the local language network and is highest amongst those with least education.

In the second part of the study, we rationalise and estimate a model of language transmission within families. We show that only about two-thirds of children with at least one parent who speaks an Indigenous language, learns to speak it. This proves that a large share of indigenous families in Mexico are at 'the margin' of the decision of whether to teach or not to teach the native language to their offspring.

The results show that the key social determinants of language transmission are the number of parents and other adults in the household that can speak the minority language, as well as the share of people in the municipality who can speak the language. Heterogeneous parents will face difficulties transmitting the minority language because the ability of a parent to

diffuse the language depends on whether the partner acts as a barrier for transmission or as a facilitator.[2] Language transmission matrices show that in households in which both parents know how to speak the Indigenous language, 72.9 percent of their children speak it, and 55.3 percent for single mothers. This number drops to 7.9 percent in two parent households in which only the mother is Indigenous bilingual to 4.7 percent when the father is the single bilingual parent.

The empirical model of intergenerational language transmission includes parental characteristics such as education level, municipality and regional characteristics. Once the model is extended to include the estimated employment benefits, which are specific to each indigenous language, we find that larger employment benefits are associated with higher transmission of the language, particularly in rural areas. The strong effect in the rural areas is consistent with the fact that a disproportionate share of indigenous people in Mexico live in rural areas and work in traditional occupations such as agriculture or crafts. Across groups, the average proportion of the indigenous population working in agriculture is 47.6%.

Overall, the results suggest that among the numerous indigenous populations of Mexico, knowing the Indigenous language allows for broader job opportunities in occupations that the Indigenous populations specialise in. As such, learning the indigenous language can be thought of as an 'insurance' against the possibility of an unsuccessful integration to the mainstream job market, where the Spanish language dominates.

The economic literature on language skills is not broad. Particular attention has been paid to the return on language skills of migrants in the developed countries, and the generic, and reasonably well identified conclusion is that immigrants have a high return on fluency in a dominant language (Dustmann 1994, Chiswick and Miller 1995, Dustmann and Fabri 2003, Bleakley and Chin 2004, Miranda and Zhu, 2013). These studies find that immigrants who are proficient in English in the UK, the US or Australia earn 5-36% more, depending on the estimation method (OLS and various instrumental variables). Dustmann and Fabri also report a positive effect on employment in the UK. The study on German fluency in Germany

[2] It is as in the model of marriage and cultural transmission in Bisin and Verdier (2000), in which transmission of culture is modelled as the result of interactions inside the family and society, where the ability of a parent to transmit his/her cultural traits crucially depends on the choice of partner which is a function of ethnic representation.

by Dustmann (1994) suggests a wage return of 7-15% using OLS with Heckman selection. On the other hand, Yao and van Ours (2015) find only modest wage effects for immigrant women and none for men in the Netherlands with respect to fluency in Dutch.

In the developing or emerging countries, a particularly well documented relationship is the economic benefit of knowing English in India, where a substantial positive return has been reported at least by Azam et al (2013) and Chakraborty and Bakshi (2016). A somewhat different angle to the same question is provided by Shastry (2012) who shows that economic areas in India that have had a lower threshold for learning English, have grown faster due to opportunities provided by globalisation and information technology.

One study that explicitly estimates employment and wage returns to bilingualism in native language and Spanish, is Chiswick, Patrinos and Hurst (2000), who find that bilingualism is a disadvantage in contrast to speaking only Spanish in Bolivia. Their study is based on an OLS estimate that does not control for ethnicity or a more precise matching of characteristics. As language and ethnicity are highly correlated, it is possible that the negative effects of bilingualism estimated in the paper are being driven by unobserved ethnicity, and the relative economic disadvantage of indigenous groups.

For our work, another relevant and interesting study is one by Munshi and Rosenzweig (2006) which shows that the choice of language of schooling in India has long-run implications for the labour market specialisation of the pupils as they grow up. In the study, they show that working class boys are disproportionately channelled to indigenous language schools which typically lead to traditional occupations, despite high returns to English language education. The suggested reason is that traditional occupations, which depend on local job networks, provide (or are perceived to provide) economic security. The results in our study can be interpreted in a similar framework: teaching the indigenous language to children may provide 'backup' job market opportunities in the traditional sector.

Another line of literature that this study contributes to is the study of ethnic enclaves. Our results resonate for example with work by Edin, Frederiksson and Åslund (2003) and Damm (2009) who find that in Sweden and Denmark, respectively, the labour market outcomes of ethnic minorities are better if they live within their own enclaves. Neither Edin et al. or Damm explicitly study language, but it is likely that the use of a minority language among

recent immigrants is one of the key factors in creating mutual understanding, job referral networks or other valuable information that affects labour outcomes.

Compared with the existing literature, one of the main contributions of this study is to show that the logic of economic returns may also apply to minority languages. Studying this issue is difficult for two reasons. Firstly, since most minority languages are in relative decline, the ex-ante view tends to be that these languages are associated with little economic benefits. Secondly, only few countries have a large enough number of minority languages for which the relevant data on language skills can be found. In this study, the census and income surveys bring the possibility to control tightly for ethnicity and overcome a negative bias that emerges from the high correlation between bilingualism and ethnicity. Further, with census data, we can estimate economic returns to minority language bilingualism for a total of 34 languages, which makes it possible to study their relationship to language transmission, another key contribution of this research.[3]

The article is structured as follows. Section 2 gives an overview of indigenous languages in Mexico. In Section 3, we present an analysis of the employment return to bilingualism, for all indigenous languages together, and separately for each language. In Section 4 we present a basic model of language choice and transmission and continue with its empirical implementation in Section 5. Section 6 concludes. The data is introduced within sections 2, 3 and 5 as appropriate.

2. Indigenous Populations of Mexico

Mexico is integrated by a rich mixture of native cultures which results in dozens of local languages that are currently spoken. These language systems stem from 11 independent language roots and were spoken by 7.36 million people in 2020.[4] While there is heterogeneity in size and geographical variation in their location, all share being concentrated in tight

[3] By separating language from the culture and estimating life outcomes of bilinguals, the recovered estimates overcome the difficulty to measure culture and its features, something that has slowed the development of economic research of culture. In this sense, core contributions of this research rest on exploiting the inclusion of the cultural identification question to the census from the year 2000 onwards in addition to the language question.

[4] Language classification from Instituto Nacional de Lenguas Indígenas.

areas which form language clusters. Our evidence supports the view that these clusters form local language and economic networks which are central to the survival of minority languages.

In 2015, little over 1 in 20 Mexicans spoke an indigenous language, yet over 1 in 5 identified as indigenous. Figures 1 and 2 provide a snapshot of the distribution of indigenous people and languages in Mexico. Based on census data, Figure 1 shows the proportion of municipal population that 'self-identifies' as indigenous. The second map, in Figure 2, shows the proportion of people who actually 'speak' an Indigenous Mexican language. There is a clear overlap between speakers and self-identified indigenous, but also a fading of speakers at the outline of the regions where most self-identified indigenous live.

Over time there has been an important decline in the representation of minority language speakers. Estimates of 1820 suggest that 60% of the population spoke a native language, by 1889 the figure was down to 38%, to 16% in 1930 and to 5.8% in 2020. When accounting for population growth, the speaking population has remained relatively constant, at least since the beginning of the twentieth century; yet, because the tightness of the language network determines the frequency of language-specific interactions, languages depend on population representation to survive. There are fears that in the absence of institutional support, especially indigenous-language schooling, an important part of the cultural heritage of these cultures can soon be forever lost.

Minority languages are concentrated to the point that where they are mostly spoken these language clusters account for an absolute majority of the local population. This changes the interpretation of minority languages, for they are nationally minoritarian but locally dominant. Across all language groups, this bimodal distribution is a recurrent feature, as displayed in the maps in the Appendix which shows the proportion of language speakers within municipalities. [5]

[5] The emergence of well-defined clusters is remarkable and draws parallels with minority languages such as Basque of Catalan in Spain, where speakers are also highly concentrated and the economic returns appear to be local (Rendon 2007 finds a 2-6 percentage increase likelihood of employment from Catalan knowledge.)

Figure 1: Population share that identifies as 'indigenous Mexican' by municipality, 2015.

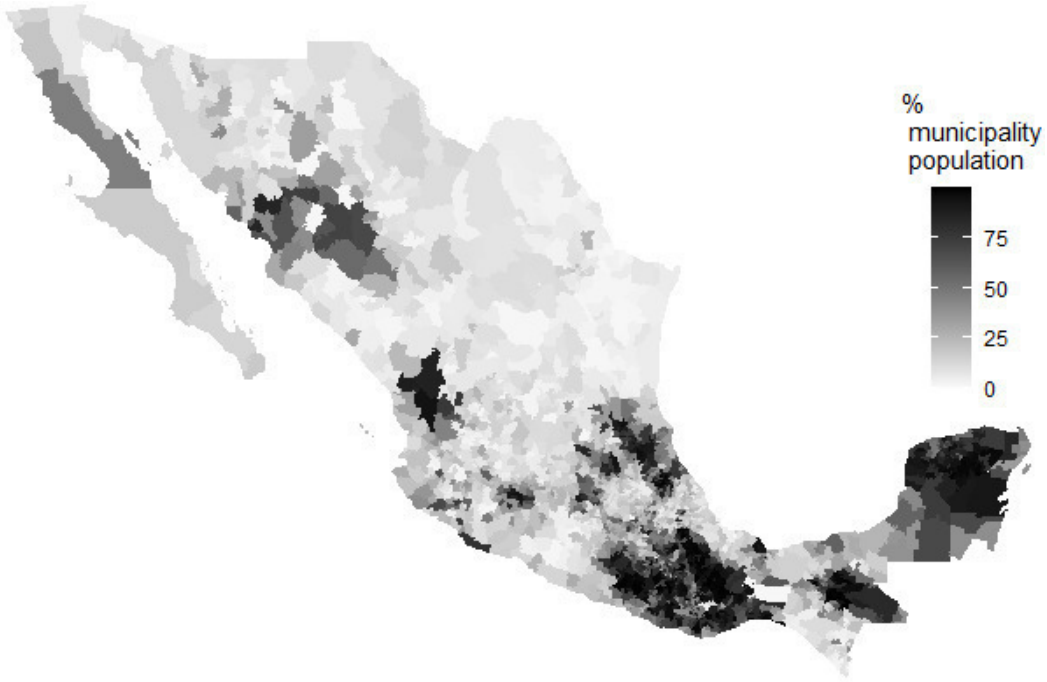


Figure 2: Population share that speaks an Indigenous Language, by municipality, 2015.

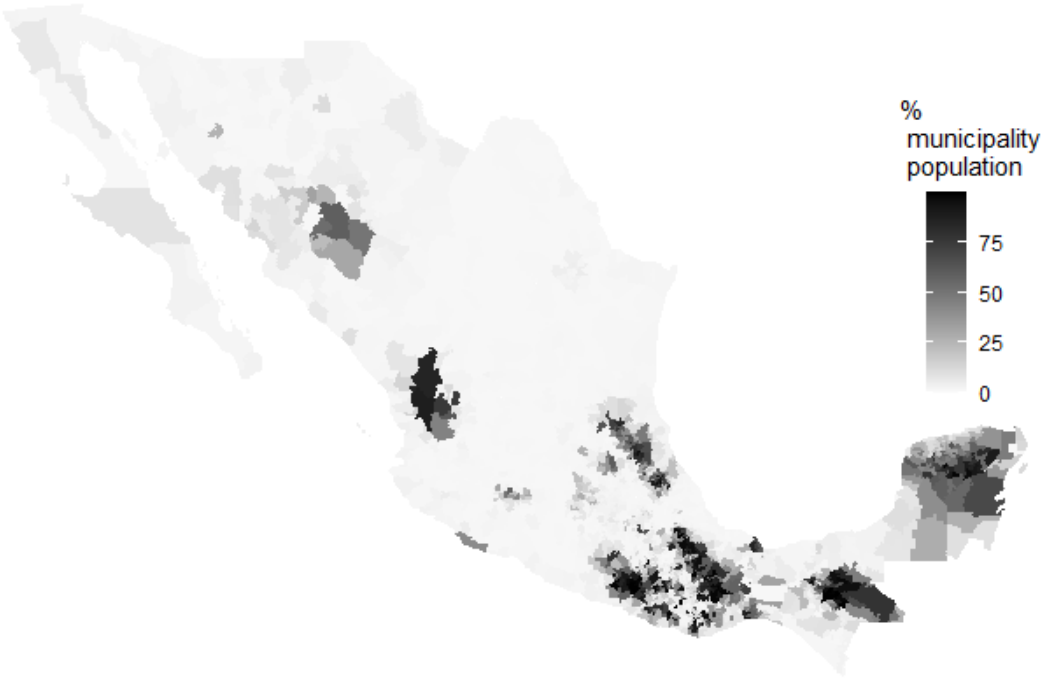


Figure 1 and Figure 2 based on own calculations from Mexican Census 2015.

Table 1 best summarizes geographical information about language groups by providing four different measures of language concentration. The language groups in the table all have more than 10 thousand speakers, and ten have a quarter of a million speakers or more. In the table, the languages (column 1) have been ordered by the total number of speakers (column 2). The two most spoken indigenous languages are Náhuatl and Maya, which together account for 36% of the minority language population (23.8% and 11.9% respectively). The third column, provides the proportion of the language population with respect to the total population. This measure shows that languages are small in relative terms as Náhuatl and Maya account for only 1.44% and 0.72% of the total population.

The fourth and fifth column bring a local dimension to these languages. The first of these provide the total number of speakers that live in a 'cluster-municipality', defined as a municipality in which either: i. over 30% of the population speak a particular language; ii. over 10% of the national language population is concentrated in that municipality, or iii. more than 10,000 same-language speakers live in the municipality.[6] Cluster municipalities are effectively municipalities with a high density of speakers and/or relatively large aggregate numbers of them. Each language has its own cluster(s) and there is little between-group mixing, although some examples of territorial overlap are discussed. At an aggregate level, of the total native speaking population, 78.22% live in areas that meet the criteria for cluster municipalities.

The fourth column gives the share of the cluster-municipalities' population that speak the minority language. This is an 'intensive-margin' measure that shows the extent of local language representation. Most language groups account for a considerable share of the local population. For example, of the 1.72 million Náhuatl speakers in 2015, 1.32 million lived in cluster-municipalities, and made up for 19.9% of these municipalities' population. For several languages, the density of speakers in the local area, measured by their proportional representation, is even much higher; such is the case of the Mixe language that, while spoken by 134 thousand, 0.11% of the national population, makes up for 68.57% of its corresponding cluster population.[7]

[6] 33 spoken languages with a total population of 73,316 have no identified cluster.

[7] With few exceptions, language groups make up single digit representations of the aggregate cluster population Smaller languages such as Chontal de Tabasco, Mayo, Huichol and Yaqui.

Table 1: Indigenous Language Clusters: Descriptive Measures of Language Concentration

Language	National		Cluster		Cluster	# Municipalities	
	Total	%	Total	%	Intensity %	Cluster	Language
Nahuatl	1,724,800	1.44	1,324,762	19.99	76.8	136	1484
Maya	861,238	0.72	824,120	21.54	95.69	113	499
Tzeltal	561,224	0.46	501,256	45.72	89.31	17	330
Mixteco	508,050	0.42	388,706	28.3	76.5	124	1016
Tzotzil	494,738	0.41	431,302	43.56	87.17	30	303
Zapoteco	464,224	0.38	332,936	50.34	71.71	141	996
Otomi	309,344	0.25	205,382	12.79	66.39	24	612
Totocana	267,868	0.22	205,218	41.9	76.61	32	578
Chol	251,942	0.2	220,394	32.31	87.47	10	240
Mazateco	240,518	0.2	172,320	61.78	71.64	28	533
Huasteco	174,434	0.14	136,172	31.01	78.06	10	307
Mazahua	151,790	0.12	103,464	16.87	68.16	7	356
Purepecha	143,360	0.11	111,860	17.26	78.02	12	297
Tlapaneco	134,592	0.11	107,364	44.57	79.76	9	233
Chinanteco	134,504	0.11	107,292	31.46	79.76	20	392
Mixe	134,404	0.11	99,142	68.57	73.76	23	495
Tarahumara	75,944	0.06	52,166	30	68.69	8	179
Zoque	67,342	0.05	44,752	33.82	66.45	10	178
Amuzgo	57,124	0.04	50,820	38.97	88.96	5	130
Tojolabal	54,316	0.04	48,170	31.59	88.68	2	66
Huichol	52,318	0.04	39,054	7.71	74.64	5	177
Chatino	51,864	0.04	46,346	57.78	89.36	9	147
Popoluca	48,974	0.03	42,178	33.02	86.12	4	118
Tepehuano	46,820	0.03	36,864	34.21	78.73	3	69
Mayo	42,270	0.03	33,442	3.97	79.11	5	71
Cora	28,472	0.02	22,618	52.88	79.43	1	50
Chontal de Tabasco	28,060	0.02	26,498	2.4	94.43	4	17
Triqui	27,490	0.02	20,390	3.67	74.17	5	135
Huave	19,924	0.01	16,894	59.77	84.79	3	93
Yaqui	19,478	0.01	11,820	7.42	60.68	1	61
Popoloca	18,012	0.01	10,924	20.83	60.64	1	65
Cuicateco	13,318	0.01	9,030	55.35	67.8	6	116
Pame	11,842	0.01	9,674	23.21	81.69	2	36
All	7,220,598	6.02	5,793,330	32.26*	78.22*		

∗ Arithmetic average across languages.

The Cluster Intensity column of Table 1 provides a measure of within-language concentration. The statistic provided is a the percentage of the overall language population that lives in its cluster area. Effectively, it corresponds to the total cluster population (column 4) over total speaking population (column 2) and shows that most of the speaking population live in their own cluster areas. So, while column 5 shows a measure of density of each language network, column 6 shows the extent to which these cluster areas represent the existence of language speakers in the country.

The last two columns of this table count the number of municipalities that make up each language cluster and the total number of municipalities in which each language is spoken. Out of the large languages, we see that Mazateco, spoken by 240 thousand people, and thus making 0.2% of the national population, is spoken in 533 municipalities, but only 28 of these are cluster municipalities, which account for 71.64% of the overall Mazateco language speakers, who themselves represent 61.78% of the cluster population.[8]

As languages are distributed in well-defined geographical, their social and economic effects must be coming almost entirely from local effects. The majority of the language groups are located at the centre and south of the country and the networks that emerge from their grouping could to be central for understanding the survival of minority languages and their varying effects.

3. Labour Market benefits of bilingualism

In this section we estimate the employment effects of attaining indigenous bilingualism conditional on being of indigenous origin. The analysis is based on two different data sources: Mexican Censuses (from 2000, 2010 and 2015) and the National Household Income and Expenditure Survey (ENIGH, from 2014, 2016 and 2018). The latter data is a more detailed employment and income survey at the cost of a smaller sample size.

We estimate labour returns to languages in two steps. Firstly, we use all the six datasets above to estimate the wage and employment likelihood return, to both indigenous languages

[8] Likewise, Mixe is spoken in 495 municipalities (of 2,456); 73.76% live in 23 of these municipalities, which are the language cluster. The remaining 26.24% Mixe speakers are in the remaining 472 municipalities.

and Spanish. The comparison groups are the matched monolinguals; in the first case the Spanish-only speaking Indigenous people and in the latter case the Indigenous-only monolinguals. Earnings returns are estimated with the survey since censuses are ridden with non-response to wages; estimates of employment returns from the census are corroborated with the income survey. Secondly, we estimate the economic returns to each indigenous language separately. This can only be done with the Census, since the ENIGH data is not large enough to capture separate returns.

The datasets contain information about ethnic group and knowledge of ethnic language. Distinguishing between ethnic group and knowledge of language is only possible from the 2000 census onwards (2010, 2015) and is central to the matching estimator that we construct. All individual variables in our analysis (gender, birthplace, employment, schooling, household composition and age), as well as locality and municipality level characteristics, such as rural-urban status, are constructed from these sources. Summary statistics for the censuses are presented in Appendix Table 6, and the corresponding summary statistics for the ENIGH data are in Appendix Table 7.

Disentangling the effect of minority language bilingualism from other socio-economic factors is based on a combination of matching and OLS estimation. In the first step, we restrict the sample tightly to only indigenous working age men who live in families in which there is at least one indigenous speaker, in a way that the indigenous language corresponds to the main indigenous language within the municipality. Because the sample is restricted to indigenous men that live in a household that speaks the dominant language of the place of residence, this matching guarantees that the control group has a minimal social distance to the treatment group. In this data we observe state of birth and residence 5 years prior to the census interview, a desirable feature because our estimates then capture the effects of bilingualism of long-term indigenous residents.

The baseline specification is summarised in the diagram below and consists of: males who are 25 to 64 years old, long term residents, self-identify indigenous, live in an indigenous speaking household and the language spoken in the household matches the dominant language in the municipality of residence. This group is then divided into treatment depending on whether they are Indigenous bilingual or control for when they are monolingual Spanish speakers. After the inclusion of all the restrictions on the data, we are left with a sample of

358,347 individuals for our estimates for 2015, 419,964 for 2010 and 136,416 for 2000. Using similar restriction in the ENIGH data, we are left with 12,188 observations from pooling the three surveys.

$$d_i = 1 \begin{cases} \text{sex = male} \\ \text{age = (25 - 64)} \\ \text{long term resident =1} \\ \text{indigenous belonging = 1} \\ \text{indigenous speaking family = 1} \\ \text{family language = mun language} \\ \text{Languages: Spanish and Indigenous} \end{cases} , d_i = 0 \begin{cases} \text{sex = male} \\ \text{age = (25 - 64)} \\ \text{long term resident = 1} \\ \text{indigenous belonging = 1} \\ \text{indigenous speaking family = 1} \\ \text{family language = mun language} \\ \text{Languages: Spanish only} \end{cases}$$

.

To recover subgroup effects, this matching sample is further tightened with the inclusion of years of schooling and language specific indicators.

$$+ \begin{cases} \text{years of schooling (0-18)} \\ \text{language (33 languages)} \end{cases}$$

.

In the second step, we estimate the likelihood of employment with a linear probability model summarized as:

$$employment_i = \gamma_0 + \lambda * d_i + \Lambda X_i + \varepsilon_i \tag{1}$$

$$wage_i = \gamma_1 + \mu * d_i + \Omega X_i + \epsilon_i \tag{2}$$

In these equations, $employment_i$ is a dichotomous variable indicating whether the person is employed (=1) or unemployed (=0) and $wage_i$ corresponds to the log of wages for those who are currently employed and receiving a positive wage. In the equations, d_i is the matching identifier between the treatment group (bilingual) and the non-treated (monolingual); X_i corresponds to a fourth-order polynomial of age, linear years of schooling, locality size controls and municipality level fixed effects.

The inclusion of municipality fixed effects together with the matching in the first step guarantees that the employment comparison between mono- and bilinguals is done within the same geographic area. The parameters λ and μ are the estimates of interest and measure the percentage point difference in the likelihood of employment for bilingual working age men

and their expected earnings differential in comparison to the control of Spanish monolinguals.

As we present the results, we display them with and without the 1st step matching to show its effect on the estimates. A priori, there's a reasonable expectation that without the 1st step matching, the control group of monolinguals would include Spanish speakers who have larger social distance to the treatment group, who are also more likely to have higher socioeconomic status. This would produce a negative bias on the estimates that the matching estimator corrects.

Results of equations 1 and 2 using Census data for 2000, 2010 and 2015 and ENIGH 2014, 2016 and 2018 are summarized in Table 2. These are the set of estimates with full set of controls which include age (linear and non-linear effects), locality size controls, schooling years and municipality fixed effects. The first four columns correspond to the employment returns of the census data; the fifth and sixth column to the pooled estimates for earnings and employment from the income and expenditure surveys.

The first row of results corresponds to the returns without the 1st step matching. Notice that employment returns from this estimation are all non-negative, but with earnings as a dependent variable the estimate for minority language bilingualism is largely negative (-9.2% lower earnings).

The second row of the table corresponds to the 2-step estimates. Now, the estimates capture the difference in the probability of employment conditional on the matching constraints, in addition to the full set of demographic controls and the municipality fixed effects. Closing the distance between groups leads to an increased employment return vis-à-vis the fixed effects estimates (2.2% increase in likelihood of employment in the pooled census matching estimate, against 0.5% with fixed effects). This is consistently observed across all estimation years with the census. The matching estimator for employment returns with the income survey is 2.1%, which is very similar to the pooled census estimates. The estimate for earnings with the matching estimator provides a major upward correction with respect to the fixed effects estimates. With the matching constraints, we find minority language bilingualism to be associated with an expected 4.9 percent increase in earnings.

The last row in Table 2 summarizes the employment returns from knowing Spanish. These

estimates are subject to a similar indigenous matching process. They differ in that the treatment group are now Spanish and Indigenous language bilinguals whereas the control group are Indigenous language monolinguals (as opposed to Spanish only monolinguals). Notice that these estimates are more in tone to the existing literature of returns to language skills that focuses on returns to migrants' knowledge of the dominant language.

Table 2: Employment Returns to Bilingualism: Increase in likelihood of Employment.

λ: employment likelihood difference.

μ: percentage earnings increase.

	λ_{2000}^{Census}	λ_{2010}^{Census}	λ_{2015}^{Census}	$\lambda_{pooled}^{Census}$	μ_{pooled}^{ENIGH}	λ_{pooled}^{ENIGH}
			Indigenous Bilingual: Fixed Effects			
Indigenous	0.017**	0.015**	0.002	0.005**	-0.0922**	0.0136**
	[0.003]	[0.001]	[0.001]	[0.001]	(0.0126)	(0.00286)
N	157,991	634,627	762,538	1,555,156	44230	49323
adj. R^2	0.120	0.085	0.114	0.090	0.357	0.057
F	157.5	763.2	2146.6	3898.2	429.8	114.2
			Indigenous Bilingual: Matching			
Indigenous	0.0354**	0.0302**	0.020**	0.022**	0.0497^{+}	0.0209**
	[0.005]	[0.002]	[0.002]	[0.002]	(0.0272)	(0.00552)
N	136,416	419,964	358,347	914,727	9836	11559
adj. R^2	0.122	0.099	0.136	0.108	0.385	0.048
F	121.8	319.9	511.4	2262.5	76.74	17.08
			Spanish Bilingual: Matching			
Spanish	0.015**	0.014**	.034**	.017**	0.273*	0.0294*
	[0.003]	[0.002]	[0.003]	[0.001]	(0.0902)	(0.0133)
N	147,439	449,784	350,617	947,840	8409	10047
adj. R^2	0.125	0.105	0.145	0.115	0.397	0.055
F	136.9	317.4	496.6	2585.17	68.76	14.71

Standard errors in parentheses. $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.001$

Matched estimates account for municipality level fixed effects.

The results show employment estimates for Spanish to be between 1.4 and 2.9 percent which is a range comparable in magnitude to the minority language return. When using the income survey, the employment probability differential estimate (2.94%) also within the range of the census estimates, which spanned between 1.4 and 3.4%, depending on the cohort studied. We found however that earnings returns to Spanish are considerable (28.9% expected earnings increase), which is consistent with the notion that dominant, official languages untap larger networks and may yield higher benefits.

The period of study is relatively short to analyse long term trajectories but there appears to be a slight downward trend in returns to Indigenous bilingualism and an upward trend in returns for Spanish. Even though censuses and income survey sweeps were conducted in different years, estimates coming from either dataset are robustly similar.

Next, this chapter documents how the returns to bilingualism vary by native group, and level of education. We also document how bilingualism features in occupational choice. Table 3 presents the employment returns to each indigenous language, ordered by magnitude of the estimated effect. Effectively, these estimates correspond to the matching estimator with full set of controls but tightened so that each language group is identified separately. In the prior estimates, employment comparisons for individuals were within language and demographic group, family type and municipality, but the estimates were interpreted as an average effect across languages. Now, we gain insight to the anatomy of the heterogeneous relationships that exist between language and employment. We estimate the employment return to 33 indigenous languages, which is the maximal number of languages we can include if we require that both the treatment and the control group must have more than 60 observations in the 2nd step estimation. Figures 5 and 6 in the Appendix show an upward relationship between language returns and the proportion of indigenous language speakers working in agriculture (last column in Table 3).

Table 3: Employment Returns to Bilingualism by Language (pooled estimates)

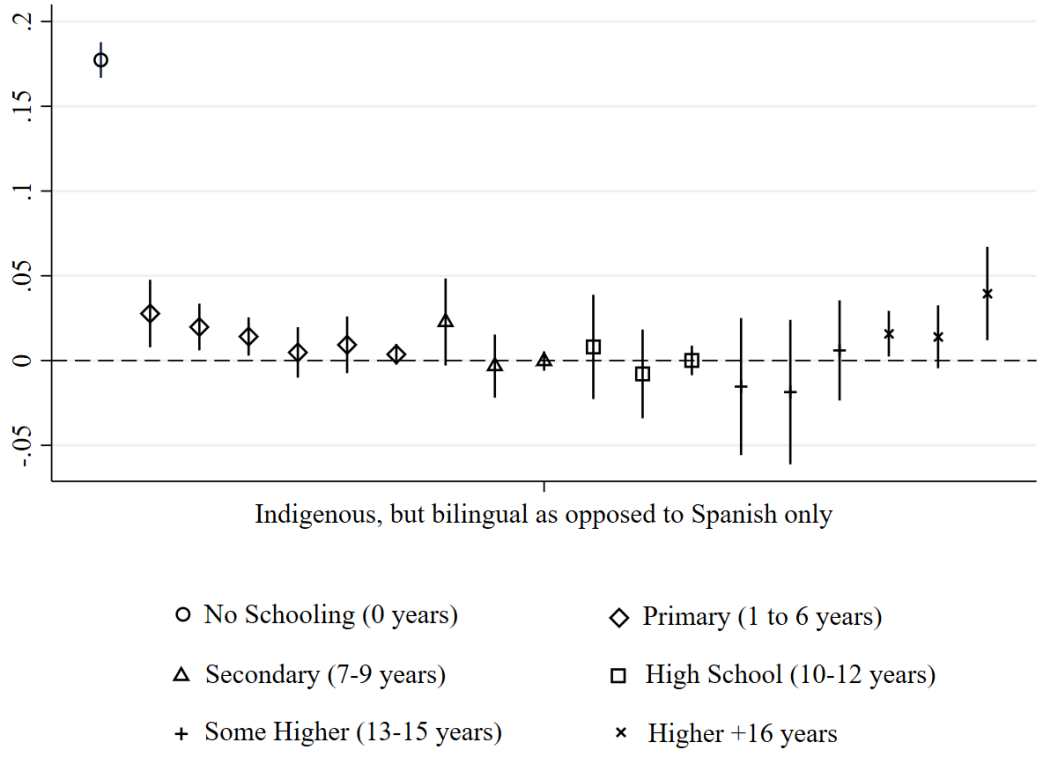
Population: 25-64 Year Old Non-Migrant Males.						
Language	λ	SE	Obs	School	Age	Agro
Maya**	0.033	0.005	139547	6.15	42.7	0.24
Zapoteco**	0.039	0.007	104327	6.51	43.0	0.33
Nahuatl**	0.020	0.005	218763	5.47	42.0	0.34
Chol**	0.078	0.020	27404	5.39	40.3	0.73
Huave**	0.119	0.038	3794	5.30	41.5	0.54
Huasteco**	0.047	0.017	18074	5.65	42.3	0.24
Mixe**	0.069	0.027	26492	5.32	41.5	0.47
Mazateco*	0.059	0.025	44340	4.62	41.1	0.48
Tzotzil*	0.058	0.024	79846	4.22	39.2	0.62
Mazahua*	0.033	0.015	9913	5.10	43.6	0.22
Chontal de Oaxaca*	0.079	0.037	1189	5.50	47.7	0.66
Huichol*	0.255	0.128	7768	5.51	39.4	0.31
Mayo+	0.037	0.019	2887	6.47	45.8	0.37
Zoque+	0.035	0.018	14319	4.44	41.6	0.65
Purepecha+	0.034	0.018	11451	6.23	41.8	0.29
Pame+	0.093	0.056	1061	3.80	41.9	0.55
Chinanteco	0.031	0.019	22346	5.39	41.8	0.64
Tojolabal	0.091	0.060	3344	3.91	39.6	0.82
Tlapaneco	0.060	0.040	25928	5.69	40.0	0.56
Popoluca	0.049	0.034	8572	4.04	41.1	0.67
Popoloca	0.070	0.055	1515	3.93	41.2	0.35
Otomi	0.012	0.011	32536	5.67	43.4	0.25
Cuicateco	0.023	0.024	5285	4.49	44.0	0.70
Tzeltal	0.018	0.019	66095	4.95	39.4	0.73
Mame	0.040	0.053	737	3.82	45.6	0.60
Cora	0.021	0.055	5361	4.42	40.7	0.48
Yaqui	0.013	0.040	1047	7.23	43.6	0.26
Chocho	0.006	0.082	197	8.40	48.2	0.46
Totocana	0.001	0.021	47017	4.99	42.7	0.49
Mixteco	-0.001	0.009	95892	5.09	42.1	0.41
Amuzgo	-0.007	0.028	11847	3.77	39.9	0.58
Chontal de Tabasco	-0.016	0.026	2140	7.63	42.4	0.27
Tarahumara	-0.017	0.026	10828	4.18	41.6	0.36
Chatino	-0.016	0.020	11584	3.59	41.6	0.60
Pop 25-64				9.8	41.3	0.10

Pop 25-64: corresponds to all non-migrant males (indigenous and non-indigenous).

Notes: ** P<0.001, * P<0.05, + P<0.1 Agro Share: share of the population working in the agricultural sector. Average School: average years of schooling.

In a further tightening of the matching estimator, the sample is partitioned by education. The employment returns for the 18 school year groups is depicted in Figure 3, where it becomes clear that the employment benefit from Indigenous bilingualism is substantially larger for those with least schooling. A close examination shows that returns are positive for individuals with less than half of primary school completed, but also for individuals at the top end of the education distribution. One potential explanation for the returns to non-schooled individuals is that for illiterates, communication is constrained to oral forms which makes languages a central asset in the social and economic life. A robustness check in which the matching sample is restricted by literacy status support this hypothesis (available on request). A robustness check in which the matching sample is restricted by literacy status support this hypothesis (available on request).

Figure 3: Indigenous language employment returns by school attainment (pooled matching 2010-2015)



Occupational statistics summarised in Tables 10, 11 and 12 in the Appendix, are connected. The first two of these tables provide statistics for both the treatment and control group.Table

10 describes the general framework of employment at a national level and tells that bilinguals are a couple of percentage points more likely to be employed but twice as likely to be independent workers, have multiple jobs, be informal workers and work without any kind of a contract.

Table 11 provides the percentage of individuals from each educational level in each occupation category; this describes the occupational distribution of men within schooling group. It becomes apparent that the ability to speak native languages comes with an increased likelihood of working in agriculture. The table effectively provides the expected occupation and shows Indigenous bilingual men are disproportionally represented in the agricultural sector across all schooling levels, and more so for the lower educated (whereas, for the whole sample, 43.1% of the Indigenous bilinguals worked in agriculture -and 15.6% of the Spanish monolinguals-, the number jumps to 60.4% for bilinguals with no schooling). "Technicians and associate professionals" is another sector that responds to education, while accounting for 3.2% of the overall indigenous language population, it absorbs over 35% of the highest educated ones.

A closer examination of occupation statistics is in Table 12 which provides occupational representation differentials across the sectors analysed in Table 11. The third and fourth column measure the proportion of jobs within each sector (column 1) taken by different occupations (column 2) for both the treatment and control group. The last column provides a ratio of these measures.

As has been documented in the previews statistics, agricultural workers (and also forestry and fishing related activities) are occupations in which the bilinguals are heavily represented. Observe also that Indigenous bilingual men are overrepresented in niche traditional occupations such as "Weavers" and "Artisans" where, conditional on sectoral choice, they are up to five times more likely to participate than their monolingual counterparts. Also, some occupations of political nature, such as "Government Officials" or "Directors of Political, Union and Civil Organizations" where, from this measure of representation, 6.3 and 2.3 times larger respectively. Perhaps in connection to the positive employment returns found for the highly schooled individuals, overrepresentation is present for workers of the educational sector. This evidence of representation is in line with the idea that language-specific labour market opportunities exist and that these networks help explain employment returns

to minority languages.

4. Transmission of language

When an individual remains monolingual, direct communication links are limited to other monolingual or bilingual speakers who share the language. A native speaker of a particular language should choose to learn another language if the utility gain derived from increasing communication links outweigh the costs of learning (a simple model of bilingualism, but not fully applicable to our study, is provided by Church and King, 1993).

In bilingual or multilingual environments and families, roughly the same idea applies to the efforts of parents to teach their children a particular language. Parents may master a menu of languages, and associate languages with different expected long term social, cultural and economic benefits. The costs of teaching a particular language to children may also vary greatly depending on the availability of speakers in the household and the exposure of the children to the language in the local environment. If the utility derived from knowing a language is increasing in the number of speakers and the costs of learning the language decrease with higher exposure to the language, cost-reducing and benefit-enlarging externalities make the efforts to learn a language an increasing function of the number of potential speakers.

Suppose that parents maximise the expected net utility of their children, and that language choice is the feature to be considered. Denote this utility as $u(language, X)$, where X accounts for the rest of the relevant things to language, the individual and the setting. Expressed as $u(language, X) = v_{lang}(network, X) - c_{lang}(network, X)$, where the utility, v, and costs, c, of learning the language depend on language network and X.[9] Let $network$ be an increasing function of the proportion of speakers in the local area, denoted by p, so that $\partial network/\partial p > 0$. then $\partial u(language, X)/\partial p > 0$, both because $\partial v(network, X)/\partial p > 0$ and $\partial c(network, X)/\partial p < 0$.

This formulation implies three possible outcomes. If $v > c$, $\forall\ p$, parents make an effort to pass their own language even in the absence of speakers in the local area. The second

[9] When the costs and benefits of learning a language are separable. This simplifying assumption is not instrumental for the argument.

equilibrium happens when $c > v \ \forall \ p$, and is then opposite; it implies that even with large number of speakers in the area, the language is not transmitted. Lastly, when there are individuals at the margin of transmitting the language, the skill is passed-on whenever the language is spoken by a sufficiently large proportion of the population, say p^*.[10] Because of network effects on c and v, insofar the actual language proportion exceeds threshold p^*, all else constant, the benefits exceed the costs of learning ($v > c$). As transmission is a function of p, this formulation places language clusters at the centre of the survival of minority languages. Notice that this results from pure externalities of language networks.

In the Mexican setting, as in most countries with minority cultures, languages lack proper institutional infrastructure and bilingual school education is underdeveloped. Institutional support and infrastructure can be incorporated to the framework under the idea that these reduce the cost of language acquisition, and hence the network threshold, p^*, lowers.

Acknowledging the role that networks play in the transmission of the languages leads to the discussion of what are the relevant features of language networks and whether different networks exist. So far, the discussion has centred around the local area network, where the density of the network, as viewed from the proportion of minority language speakers, has incidence on the likelihood of transmission. The focus next is on the family and employment, where a similar logic apply.

In the family network the core idea is that the costs of teaching a language to a child declines with more adults in the household who are able to speak it. Furthermore, the existence of relatives who speak an indigenous language also increase the social benefits of knowing the language. In the estimates below, we show that the transmission is higher when both parents can speak the minority language, than in mixed couples where only one of the parents can speak it. Additional extended family members in the household who speak the minority language also increase the likelihood of transmission.

Commanding an additional language is a skill, with a potential positive economic return. In the context of Mexico, the economic benefits from native languages are likely to exist due to employment networks in certain professions, such as in agriculture, traditional in-

[10] At $p = p^*, u(network, X) = v(network, X) - c(network, X) = 0$

dustries and professionals in education. Our main hypothesis is that the decision to pass the language is partly informed by the perceived economic opportunities that the language skill may provide to their children. One can assume them to be informed by the existing economic returns that the parents' generation has enjoyed. Overall, we build a simple model of intergenerational language transmission as a function of the language networks and the economic benefits. In the formulation of the problem, the network of the language speaker is a function of the family language structure and local area language characteristics, as in $network(family, local, economic)$

$$P_i(\text{Language Passed}|X) = F(\underbrace{network}_{Family,Local,Economic}, \underbrace{Z}_{Controls}) + \epsilon_i \tag{3}$$

Specifically, *Family* correspond to language resources in the family; *Local* to language resources in the local community, *Economic* to the employment return to the language and Z to controls of the family, language and municipality.

The study focuses on families with both parents present. In such cases, the language resource in the family will be measured with variables that consist primarily on whether both or only one of the parents can speak the native language. We also take into account whether there are other adults in the household who can speak the language (such as grandparents). We assume (and test) that each additional adult who can speak a native language in the household, generally increases the likelihood that the language is passed down to the next generation.

For local elements, the main focus is on the strength of the language in the local area, for which we compute the proportion of people in the municipality who speak the same minority language as the household does. In practice, this measures the potential interactions that can be made using the minority language in the local area. In our data of bilingual families, the average family is located in a municipality where 52% of the local population can speak the same language as the family. This confirms that the typical bilingual family lives in a 'core' of the minority languages. On the other hand, this measure has a large variation, showing that the strength of the local language network cannot in general be taken for granted, and it is important to control for it.

4.1. Data and sample

The data is based on the Mexican Census of 2015. The sample is limited to the household respondent and his/her spouse and children. Families which speak only Spanish are excluded, so that at least one of the parents states that they can speak a native Mexican language. To simplify analysis, single-parent families and families where parents speak two different native languages are excluded. As such, each bilingual nuclear family is categorised to belonging to one of the native Mexican language groups. Further, the age of the mother has been restricted to range 25-54.

Table 13 in the Appendix presents the summary statistics on the 2015 sample of households. Within the sample, 64.5 percent of parents have passed the minority language to their children.[11] In 9 percent of the families, only mother can speak the native language, and in 12 percent, only father. This implies that in 79 percent of the families, both parents state that they can speak a native language. We have not documented the Spanish skills, since it is increasingly rare that people in Mexico can't speak any Spanish. All children are exposed to Spanish by the school system.

A noteworthy fact is that only about two-thirds of children with indigenous-speaking parents learn the indigenous language (Table 6). This goes to show that a large fraction of families is likely to be 'in the margin' of deciding whether to pass the indigenous language to the next generation.

Table 13 also lists a number of key household variables that may affect the transmission of language within the household. The table includes variables at the native language group level and the municipality level. At language group level, the main variable of interest is the group-specific employment return to bilingualism, or the estimated increase in likelihood of employment from being able to speak the native language in addition to Spanish. Other variables that proxy the economic importance of the group are the group size, as well as the average wealth and education in the group. All of these variables have substantial variation across the 34 groups covered by the sample.

[11] In 93% of the families, either all or none of the children learn the minority language. Therefore we have rounded the share of children who speak the language to either 0 or 1. The language skills of children under 4 years are not defined

4.2. Results

Table 4 shows results on a number of estimations for the determination of language transmission, using family and municipality characteristics, and the economic return to languages.

The first column is the benchmark model for language transmission, and it uses only the household characteristics, as well as regional fixed effects. The first important result is that if either mother or father can't speak the native language, it is much more likely that the language is not transmitted to children. If father doesn't speak the language, the likelihood of transmission falls by 44 percentage points. Mother's ability to speak the minority language is estimated to be somewhat more important than father's (47 percentage points), which is consistent with mothers spending more time with their children than the father.

Additional adults in the household that can speak the minority language increase the likelihood of language transmission by about 4.5 percentage points per person. While the effect is statistically quite significant, the size of the effect on children is only about 1/10 of the effect of a parent's language.

With regards to the local network of minority language speakers, the first column of Table 4 suggests that if the local proportion of minority language speakers increases by 10 percentage points, the likelihood of transmitting the language in the household increases by 4.4 percentage points, which is not far from the effect that one additional adult speaker in the household has. This is a variable that has substantial variability across households, with a standard deviation of 0.3. This implies that moving a bilingual family to a municipality with 1 SD larger share of minority speakers would imply a 13.2 percentage points (0.3*0.44) larger likelihood that the language is passed to the next generation.

Parental education, age and household wealth (based on an index of items) all have a negative and significant association on the likelihood of language transmission. Of these, it is worth noting that each year of maternal education reduces the likelihood of the language transmission by about 0.7 percentage points, and one standard deviation of household wealth by about 5 percentage points. An explanation for these effects could be that further study, typically conducted in Spanish, gears the parents to overlook the potential value of the minority languages. The effect of wealth and education suggest that in general the indigenous

languages are strongly associated with lower socio-economic status in Mexico.

Columns 2 adds the employment return of the language-specific bilingualism into the model. The effect of the economic return in itself suggests a positive and statistically significant effect. Here it is important to note that since this variable varies by the 34 native groups, the standard errors are clustered by these groups. Column 3 further adds controls for the municipal level of economic deprivation. Since there are nearly 2000 municipalities covered by the sample, adding these controls allows us to address for sources of potential omitted variable bias in the model. Other variables included are indices for educational, health, housing and food deprivation. Remarkably, these have very little effect on the results of interest, suggesting that local levels of economic development are well controlled for and do not bias the results.[12]

The fact that a very large proportion of the indigenous males work in agriculture suggests that the employment return to the native languages must be partly driven by employment dynamics in this sector. If that is the case, it is possible that the families in rural areas respond to this economic benefit more than in urban areas, by making sure their children learn the indigenous language. Information from the language transmission matrices in Table 16 and 17 indicate so.[13]

This is why, in the final column of Table 4, we have interacted the employment return with urban location. The results show that in rural areas, the higher employment return is associated with an increased likelihood to pass the language, at 5% statistical significance level, whereas in the urban areas, the effect is very close to zero. In rural areas, the size of

[12] An alternative to municipal multidimensional deprivation would be to use municipal fixed effects. The problem with this approach is that since the native groups are highly regional (see the appendix maps), municipal fixed effects would not have sufficient variation in most of the country, but would instead be based on the largest cities which host multiple indigenous groups, but with few, and very selected individuals. This would not give the representative estimates we are looking for.

[13] These tables arrange households by the language-composition of parents (and occupation) and provide the share of children who are bilingual in each of household groups. Estimates of Table 16 show that in bilingual households, in which the 'head' of the household works in agriculture, 84.1% of children are minority language bilingual too, a transmission rate well above that of any other sector. Table 17 corroborates this by showing that across the ten largest spoken languages, language transmission rates are consistently higher in agriculture households.

the effect is not trivial: If the employment return to bilingualism increases by 2 standard deviations (2*.033), the likelihood of passing the language increases by 3 percentage points (2*.033*.432 = .0304).

Table 4: Factors behind language transmission: a focus on the family, locality and employment returns.

Dependent variable: Children can speak native language				
	1	2	3	4
Employment Return		.364**	.368**	.432**
		[.124]	[.125]	[.122]
Urban*Employment Return				-0.373
				[.303]
Urban				-.0401*
				[.0156]
Only mother speaks native	-.443**	-.442**	-.442**	-.439**
	[.0326]	[.0325]	[.0323]	[.0331]
Only father speaks native	-.474**	-.473**	-.472**	-.47**
	[.0341]	[.034]	[.034]	[.0349]
# Other adults speak native	.0449**	.0455**	.0455**	.046**
	[.00541]	[.00528]	[.00526]	[.00539]
% municipality share HH lang	.436**	.436**	.444**	.442**
	[.0434]	[.0433]	[.0425]	[.0428]
Mother's years of educ	-.00722**	-.00727**	-.00747**	-.00776**
	[.000689]	[.000682]	[.000714]	[.00071]
Father's years of educ	-.004**	-.00407**	-.00433**	-.00431**
	[.00059]	[.000599]	[.000642]	[.000627]
Mother's age	0.000396	0.000395	0.000363	0.000372
	[.000321]	[.000324]	[.000326]	[.000351]
Father's age	-.000594**	-.000603**	-.000638**	-.000718**
	[.000201]	[.000199]	[.000194]	[.000183]
Normalised HH wealth	-.0519**	-.0515**	-.051**	-.0422**
	[.00489]	[.0047]	[.00441]	[.00345]
Municipal controls:				
Educational deprivation			-.00141*	-.00165**
			[.000564]	[.000593]
Health deprivation			-0.0000531	0.000218
			[.000651]	[.000563]
Housing deprivation			0.000402	0.000392
			[.000324]	[.00031]
Food deprivation			.000564**	.000617**
			[.000197]	[.000211]
Constant	.581**	.571**	.59**	.611**
	[.0414]	[.0426]	[.0405]	[.0411]
Observations	227,076	227,076	227,076	227,076
R-squared	0.524	0.524	0.525	0.527

Notes: Linear probability. If at least 50% of children speak native, the family is coded as 1 in the dependent variable. Standard errors clustered at language group level, ** p<0.01, * p<0.05, + p<0.1. Employment return: language specific increased likelihood of employment. Models include regional FE.

4.3. Robustness checks

An obvious concern that arises from the estimates of Table 4 is that the result on the employment return observed could in fact be reflecting the generic socioeconomic status of the language. It may be that estimates of the return to bilingualism are not actually returns on skills per se, but that they are signals of the relative prestige of the language: Workers who belong to a higher status group, are more likely to find work (leading to the variability in employment return across groups) and also more likely to pass the language to their offspring as a 'signal' of the group membership.

Due to this concern, in Table 5 we report estimates with the inclusion of other variables that capture whether the status of the group change the estimates. If the generic group status is a source of omitted variable bias, inclusion of these variables should reduce the estimate on how much the return to bilingualism affects language transmission. The columns 1-3 of the table include, consecutively, (1) the size of the group as measured by the logarithm of the number of households in Mexico where the language is spoken, (2) The average wealth index of the households of the group and (3) the average education of the households in this group. Further, in column 4, all of these variables are included at the same time.

Table 5: Robustness check: Language group status

Dependent: Children can speak native language				
	[1]	[2]	[3]	[4]
Employment Return	.301*	.364**	.433**	.335*
	[.13]	[.131]	[.117]	[.127]
Ln Group Size	-.0098*			-.00827+
	[.00447]			[.00408]
Group avg wealth		-0.0259		-0.00341
		[.0165]		[.0231]
Group avg education			-0.0148	-0.00543
			[.00925]	[.0127]
Municipality controls	yes	yes	yes	yes
Household controls	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes
Observations	227,076	227,076	227,076	227,076
R-squared	0.525	0.525	0.525	0.525

Notes: All models include the same controls as in column 3 of Table 4. Standard errors clustered at language group level, ** p<0.01, * p<0.05, + p<0.1.

Remarkably, the results show that the effect of employment return on the language transmission appears to be fairly orthogonal to these variables. From this, we can conclude that the relative socioeconomic status of the languages is not driving the main result in Table 4.[14]

5. Conclusions

This study demonstrates two linked results on the economics of language skills, that are new to the existing literature. The study is based on data from Mexican Censuses and the National Household Income and Expenditure Survey.

Firstly, we estimate the economic return to being bilingual for 34 indigenous Mexican languages. We show that on average, observationally identical indigenous Mexican males are more likely to be employed if they can speak both the indigenous language and Spanish as opposed to Spanish only. The employment benefit varies by language and can be explained by the domination of agriculture by the indigenous groups. The result can be demonstrated with two different data sets, and emerges as we take the analysis from an OLS with municipality fixed effects to a detailed matching estimator, suggesting that unobserved omitted factors relating to socioeconomic circumstances are likely to bias the return to indigenous languages downwards in the OLS. Economic benefits of minority languages have not been shown in the literature to this extent before due to data limitations.

Secondly, we show that the employment returns to a language affect the likelihood that parents transmit the language skills to their children. We build a detailed picture of bilingual households in Mexico, and control for all key factors in the family and the local environment that affect the transmission of indigenous languages either by reducing the cost or increasing the benefits of knowing them. The result is driven by rural areas where indigenous populations dominate niche sectors such as agriculture. This result contributes to literatures on intergenerational transmission, identity formation and ethnic enclaves. For example, the results suggest that the economic benefits of migrant enclaves observed in other literature are likely to be mediated by a common language. The study also provides a unique sys-

[14] An additional measure of group status we used is whether some parts of the municipality has autonomous indigenous governance that allows for native language schooling. This had only a minor effect on the estimates, and the own effect of this variable was positive but not statistically significant.

tematic documentation on how economic factors can affect the continuation and survival of minority languages that lack the support of official institutions in developing countries. It is furthermore apparent from the results that additional language skills can be thought of as forms of insurance that allow the speakers to access niche labour markets, resonating with early results such as Munshi and Rosenzweig (2006).

These findings point to a number possibilities for future research on the economic returns of minority languages in other countries and contexts, and how policies and institutions interact with this relationship. A precondition for such research is that the censuses or household surveys document the languages spoken by individuals, and that the country in question has rich variability in minority languages. Institutional changes in language policies and labour markets can aide the identification of the effects of interest. In our study period for Mexico, there were no significant changes in the institutional setting and the formal support for the minority languages was weak.

References

[1] ˜Azam, Mehtabul, Aimee Chin, and Nishith Prakash (2013). *The Returns to English-Language Skills in India.* Economic Development and Cultural Change, 61(2): 335-367.

[2] ˜Bleakley, Hoyt and Aimee Chin (2004). *Language Skills and Earnings: Evidence from Childhood immigrants.* The Review of Economics and Statistics, 86(2): 481-496.

[3] ˜Chakraborty, Tanika, Shilpi Kapur Bakshi (2016). *English language premium: Evidence from a policy experiment in India.* Economics of Education Review, 50, Feb 2016, pp. 1-16.

[4] ˜Chiswick, Barry R. and Paul W. Miller (1995). *The Endogeneity Between Language and Earnings: International Analyses.* Journal of Labor Economics, 13(2): 246-288.

[5] ˜Chiswick, Barry R., Harry A. Patrinos and Michael E. Hurst (2000). *Indigenous Language Skills and the Labor Market in a Developing Economy: Bolivia.* Economic Development and Cultural Change, 48(2): 349-367.

[6] ˜Church, Jeffrey and Ian King (1993). *Bilingualism and Network Externalities.* The Canadian Journal f Economics, 26(2): 337-345.

[7] ˜Damm, A. P. (2009). *Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence.* Journal of Labor Economics, 27(2): 281-314.

[8] ˜Dustmann, Christian (1994). *Speaking fluency, writing fluency and earnings of migrants.* Journal of Population Economics, 7(2): 133-156.

[9] ˜Dustmann, Christian and Francesca Fabbri (2003). *Language Proficiency and Labour Market Performance of Immigrants in the UK.* The Economic Journal, 113(489):695-717

[10] ˜Edin, Per-Anders, Peter Fredriksson and Olof Åslund (2003). *Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment.* The Quarterly Journal of Economics, 118(1): 329–357.

[11] ˜García Cubas, A. (1904). *El libro de mis recuerdos.* Imprenta de Arturo García Cubas, Hermanos Sucesores, México.

[12] ˜García Segura, Sonia (2004). *De la Educación Indígena a la Educación Bilingüe Intercultural (La comunidad p'urhépecha, Michoacán, México).* Investigación temática - educación intercultural. 9(22), p. 61-81.

[13] ˜Instituto Nacional de Estadística, Geografía e Informática. *Household Census Microdata (2000, 2010, 2015).* Mexico. Extracted from: Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.3 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D020.V7.3

[14] ˜Instituto Nacional de Estadística, Geografía e Informática. *Encuesta Nacional de Ingreso y Gasto de los Hogares (ENIGH 2014, 2016 AND 2018).* Extracted from: INEGI Public Use Microdata Series [dataset]. Mexico. https://en.www.inegi.org.mx/datos/

[15] ˜Instituto Nacional de las Lenguas Indígenas (2008). *Catálogo de las Lenguas Indígenas Nacionales. Variantes Lingüísticas de México con sus autodenominaciones y referencias geoestadísticas.*

[16] ˜Jiménez-Naranjo, Y. and Mendoza-Zuany, R. (2016). *La educación indígena en México: una evaluación de política pública integral, cualitativa y participativa.* LiminaR, 14(1), 60-72.

[17] ˜Miranda, Alfonso and Yu Zhu (2013). *English deficiency and the native-immigrant wage gap.* Economics Letters, 118(1): 38-41.

[18] ˜Munshi, Kaivan and Rosenzweig, Mark (2006). *Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy.* American Economic Review, 96(4): 1225-1252.

[19] ˜Nettle, Daniel, and Suzanne Romaine (2002). *Vanishing Voices: The Extinction of the World's Languages.* Oxford University Press, USA.

[20] ˜Recondo, David (2007). *La política del gatopardo. Multiculturalismo y democracia en Oaxaca.* Capítulo 2. El surgimiento de un actor político indígena México, Publicaciones de la Casa Chata. p. 83-120. https://books.openedition.org/cemca/2066?lang=es

[21] Rendon, Silvio (2007) *The Catalan Premium: Language and Employment in Catalunya.* Journal of Population Economics. 20:669-686.

[22] ˜Salmerón Castro, F. and Porras Delgado, R. (2010). *Los Grandes Problemas de México.* Part 4, Chapter 17,"La educación indígena: fundamentos teóricos y propuestas de política pública". El Colegio de México, Vol VII, p. 509-546.

[23] ˜Sarmiento Silva, Sergio (1985). *El Consejo Nacional de Pueblos Indígenas y la política indigenista.* Revista Mexicana de Sociología. Universidad Nacional Autónoma de México 47(3) p.197-215.

[24] ˜Schmelkes, Sylvia, (2006). *La interculturalidad en la educación básica.* Proyecto Regional de Educación para América Latina y el Caribe, UNESCO. Num. 3. p. 120-127.

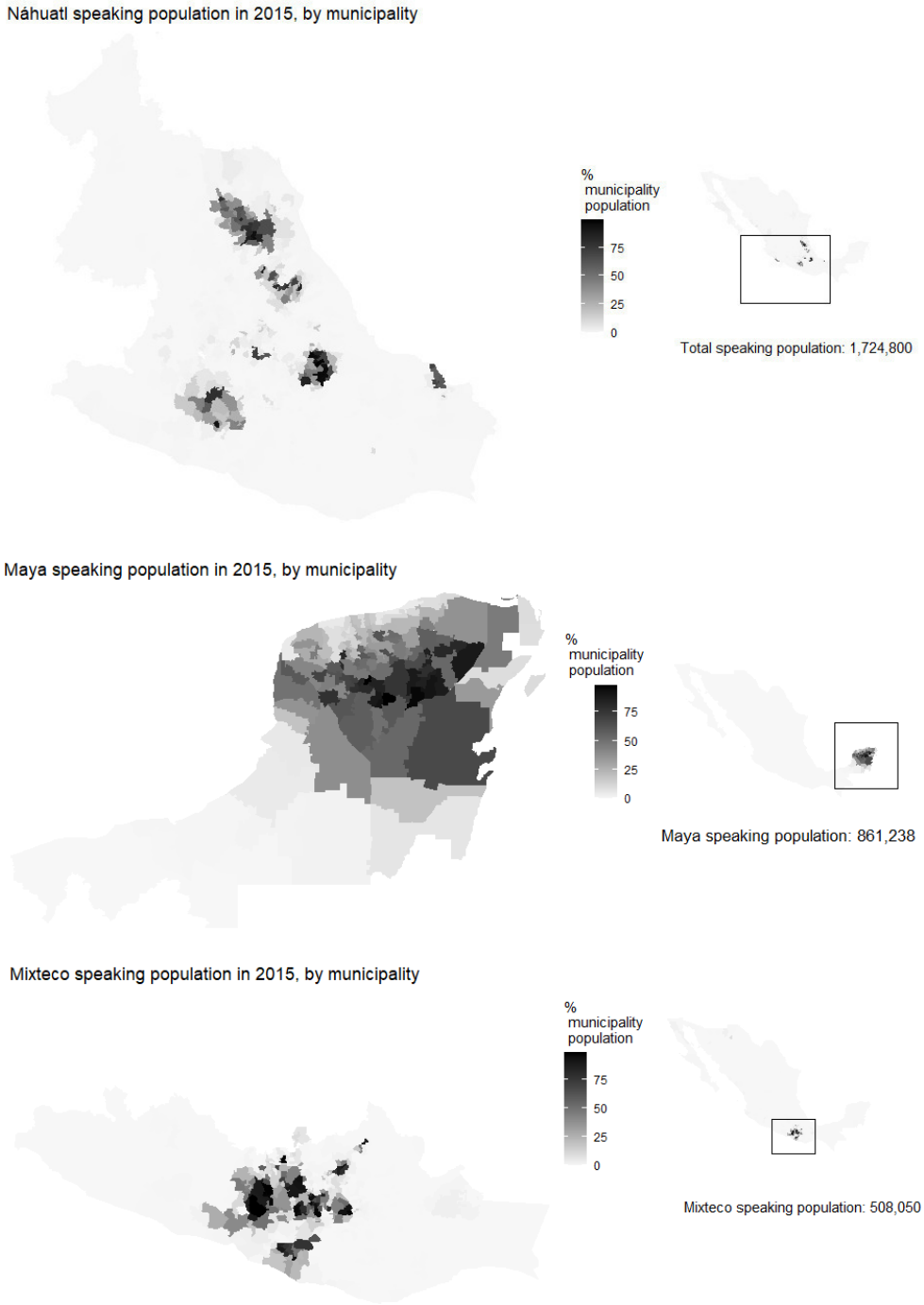
[25] ˜Shatry, Gauri Kartini (2012). *Human Capital Response to Globalization. Education and Information Technology in India.* The Journal of Human Resources, 47(2):287-330

[26] ˜Stavenhaven, Rodolfo (1988). *Anuario de etnología y antropología social.* Chapter 7: Los derechos humanos de los pueblos indios. Vol 1, México, D.F., p. 130 - 136.

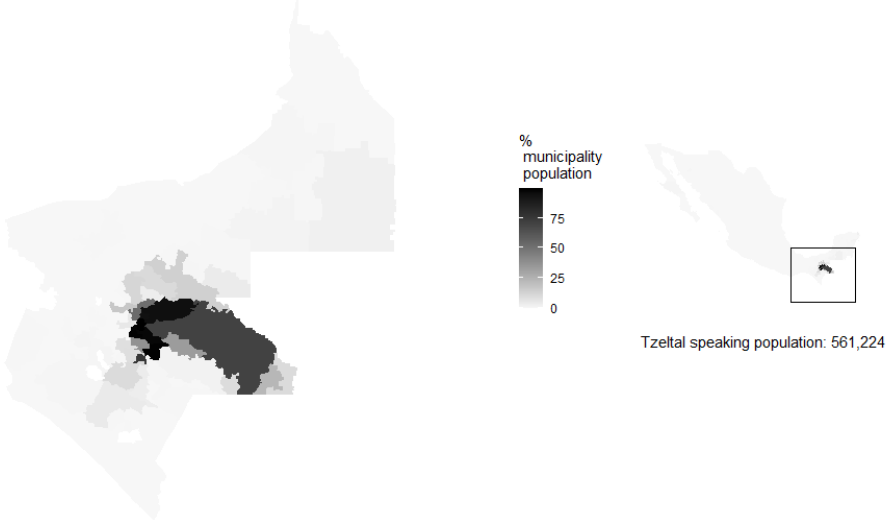
[27] ˜Yao, Yuksin and Jan C. van Ours (2015). *Language skills and the labor market performance of immigrants in the Netherlands.* Labour Economics, 34(June 2015): 76-85.

Appendix 1: Tables and Figures

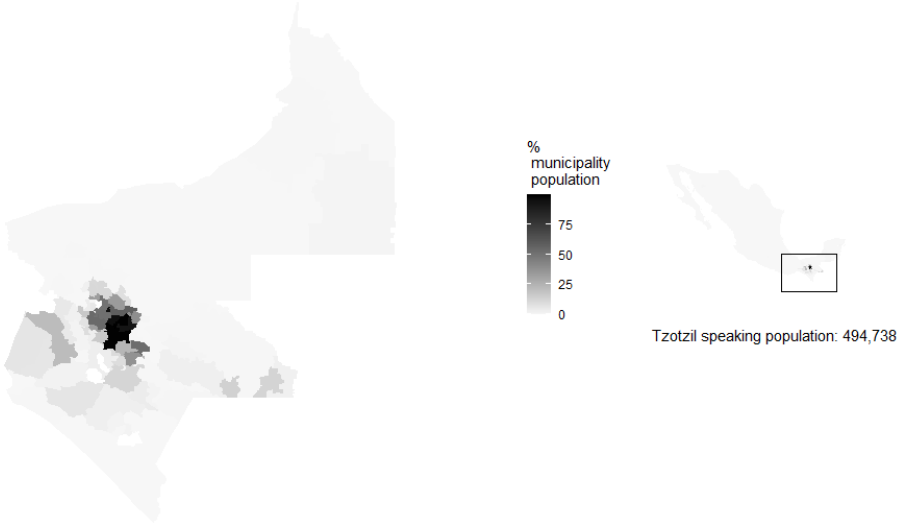
Figure 4: Anatomy of the Geographical Distribution of Indigenous Language Speakers



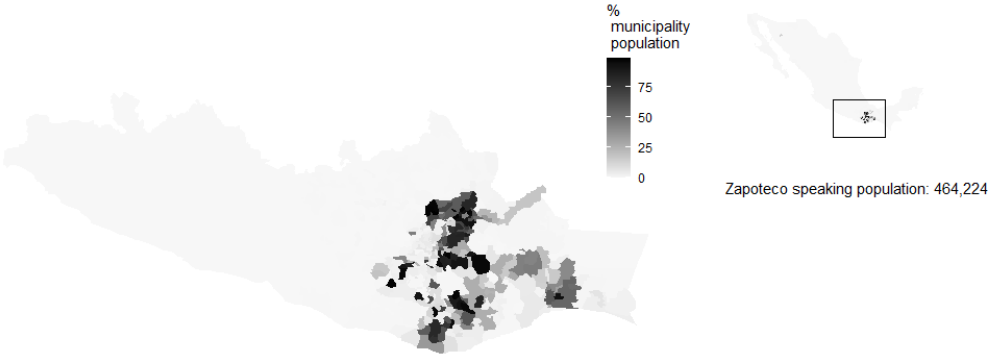
Tzeltal speaking population in 2015, by municipality



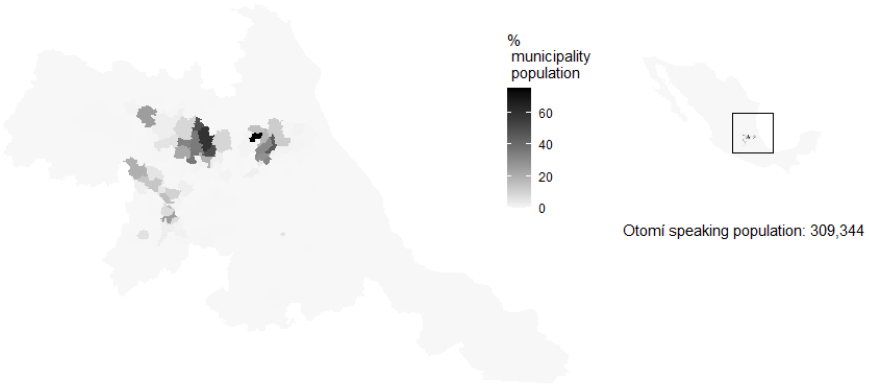
Tzotzil speaking population in 2015, by municipality



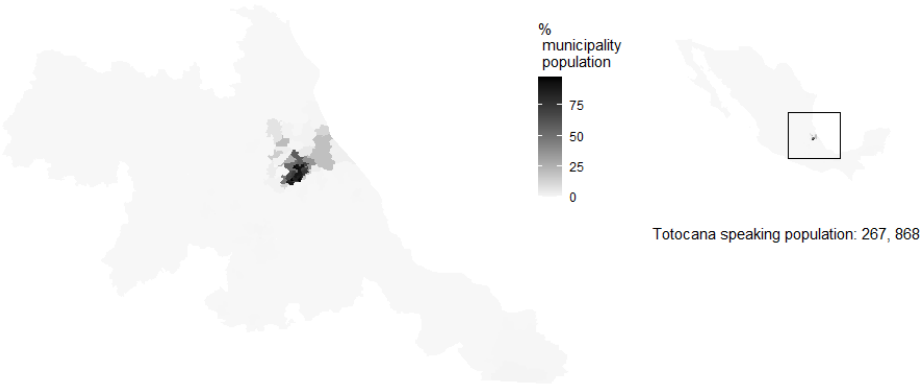
Zapoteco speaking population in 2015, by municipality



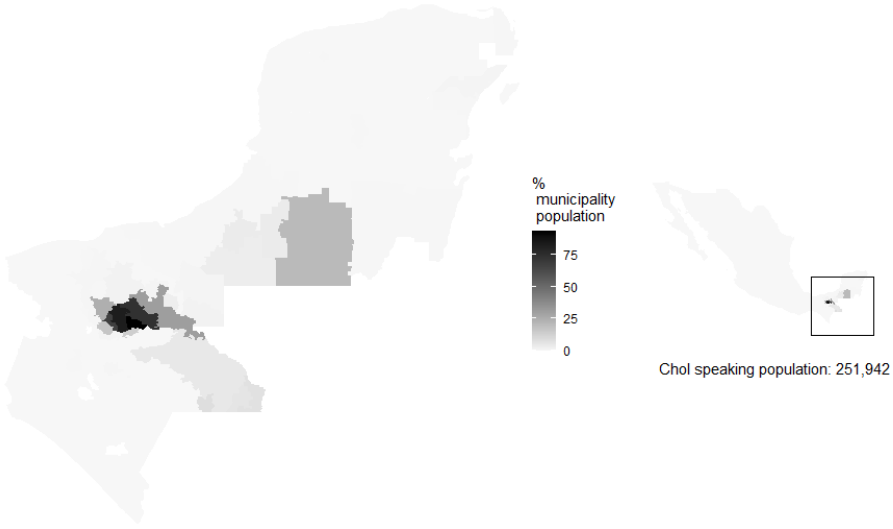
Otomi speaking population in 2015, by municipality



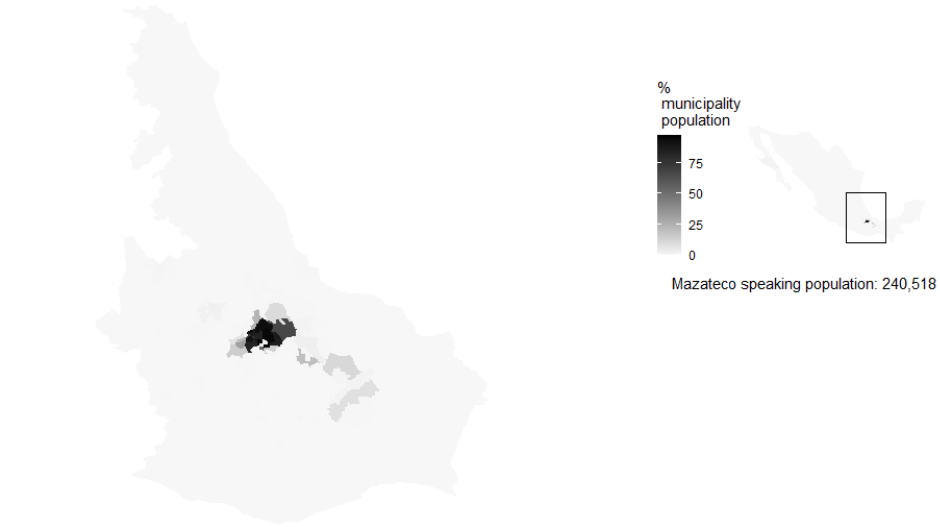
Totocana speaking population in 2015, by municipality



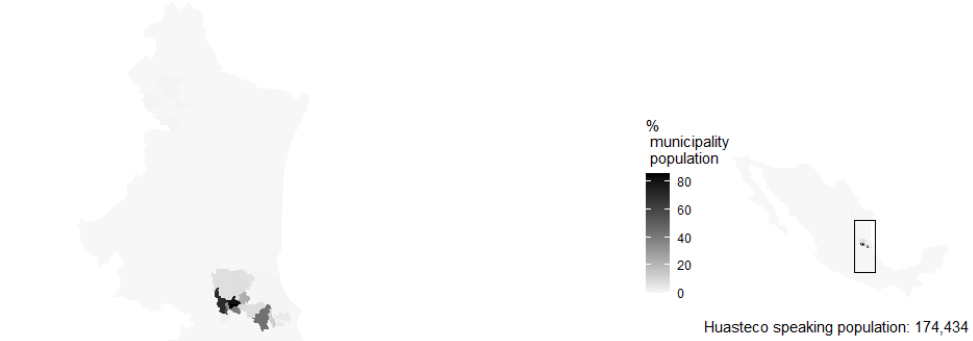
Chol speaking population in 2015, by municipality



Mazateco speaking population in 2015, by municipality



Huasteco speaking population in 2015, by municipality



Mazahua speaking population in 2015, by municipality

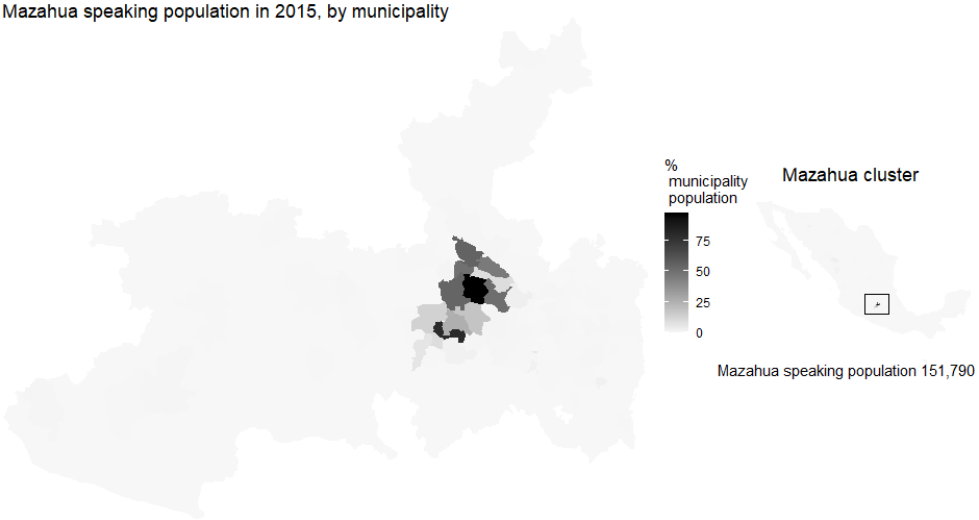


Figure 5: Employment Returns and Agricultural Work Shares, by Language (languages with statistically significant employment returns)

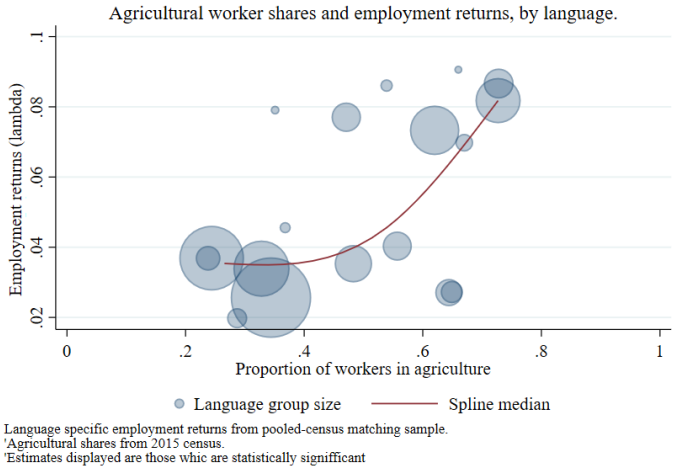
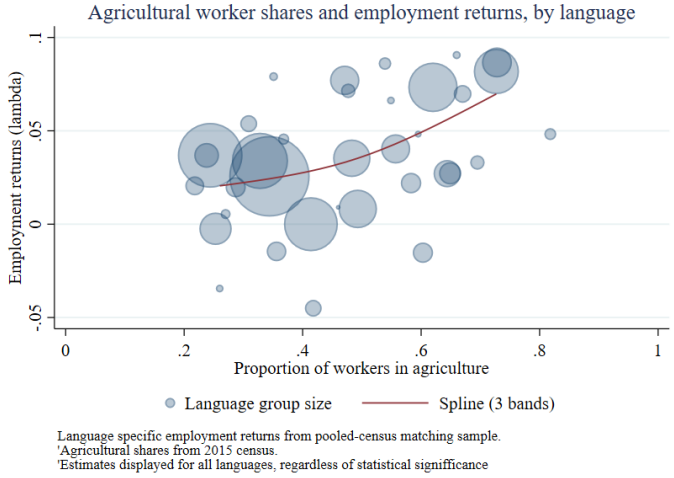


Figure 6: Employment Returns and Agricultural Work Shares, by Language (all languages)



.
.
.
.
.

Table 6: Summary Statistics Census 2015 by population groups

Indigenous	Speaks Indigenous	Average Age	Illiteracy Rate	School Years	Work Status	Sample Size
All non-migrant male, 25-64						
0.21	0.07	41.4	0.04	9.60	0.85	2,352,008
All non-migrant male, 25-64 Identified with ethnic/indigenous group.						
1	0.28	41.6	0.07	8.03	0.83	799,977
All non-migrant male, 25-64 Identified with ethnic/indigenous group Speaks Indigenous Language.						
1	1	42.3	0.16	6.02	0.80	357,369

* Estimates for three different samples. Non explicit variables as follow:
Indigenous: Share of population that self-identifies as indigenous; Schooling: Average years of schooling; Illiteracy Rate: share that "does not know how to read or write a message"; Work Status: share of the population working

.
.
.
.
.
.

Table 7: Summary Statistics ENIGH 2016, by population groups

Indigenous	Speaks Indigenous	Average Age	Illiteracy Rate	Post Primary	Work Status	Sample Size
All non-migrant male, 25-64						
0.30	0.06	41.9	0.04	0.29	0.91	56,217
All non-migrant male, 25-64 Identified with ethnic/indigenous group.						
1	0.20	42.1	0.07	0.39	0.92	17,776
All non-migrant male, 25-64 Identified with ethnic/indigenous group Speaks Indigenous Language.						
1	1	42.9	0.15	0.60	0.94	3,968

* Estimates for three different samples. Non explicit variables as follow:
Indigenous: Share of population that self-identifies as indigenous; Schooling: Average years of schooling; Illiteracy Rate: share that "does not know how to read or write a message"; Work Status: share of the population working

Table 8: Employment Statistics for Indigenous Populations (Matching Sample)

	Employment Statistics	
	Indigenous Bilingual	Spanish Monolingual
Employed	.943	.917
Independent Workers	.447	.244
Multiple Jobs	.201	.103
Formal workers[1]	.159	.344
Contract[2]	.156	.318
Weekly Hours of Work	45.3	48.9

[1] Defined by whether they have access to social security and health (IMSS, ISSTE, Pemex, Military)

[2] Working under a contract. ENIGH 2016, matching estimator sample.

Table 9: Occupation choice by bilingualism, percentages by level of education

Industry Occupation for Indigenous by Bilingual Status (and Schooling Years).
Spanish monolinguals (d=0), Indigenous bilinguals (d=1).

	Schooling					
	0	1-6	7-9	10-12	12+	All
	d=0					
Industry-Occupation						
Legislators, senior officials and managers	0.1	0.3	0.7	2.5	8.5	2.2
Professionals	0.3	0.5	0.7	1.2	30.5	5.7
Technicians and associate professionals	0.4	0.5	1.4	5.4	14.1	4
Clerks	0.4	0.8	2.3	6.4	7.5	3.4
Service workers and shop and market sales	8.7	10.8	18.1	25.5	15.2	16.5
Skilled agricultural and fishery workers	40	28.6	13.2	6.3	2.2	15.6
Crafts and related trades workers	23.8	28.4	27	22.5	11.7	23.9
Plant and machine operators and assemblers	5.4	11.6	20.1	18.1	6.6	14.5
Elementary occupations	21.3	18.9	16.5	11.8	3.9	14.4
Armed forces	0.1	0.1	0.5	0.9	0.3	0.4

	Schooling					
	0	1-6	7-9	10-12	12+	All
	d=1					
Industry-Occupation						
Legislators, senior officials and managers	0.1	0.2	0.4	1.2	5.4	0.6
Professionals	0.2	0.3	0.6	0.9	21.5	1.7
Technicians and associate professionals	0.2	0.4	1	6.4	35.1	3.2
Clerks	0.1	0.3	1.1	4.1	5.7	1.2
Service workers and shop and market sales	4.7	6.9	14.8	21.6	9.4	10.1
Skilled agricultural and fishery workers	60.4	51.7	34.8	23.2	6.6	43.1
Crafts and related trades workers	14.3	17.9	20.2	16.6	7.8	17.3
Plant and machine operators and assemblers	2.2	4.4	8.5	11.2	4	5.7
Elementary occupations	18.2	18.3	18.7	14.1	4.7	17.1
Armed forces	0.0	0.1	0.6	1.3	0.3	0.3

Note: Males 25-64 by ethnicity and language domain. Observations
with non-identified professions excluded when computing this set of statistics.

Table 10: Occupation within Industry: Differentials of Ethnic Indigenous by Language Status

Industry	Occupation within industry	Within Industry Share d=0	Within Industry Share d=1	Relative difference
Legislators, senior officials and managers	Directors and managers in health, educational, and social services	9.8	14.7	1.52
	Coordinators in health, educational, and social services, and grading judges[1]	7.6	11.6	1.52
	Directors and managers of museums, cinemas, and other establishments	1.8	3.3	1.85
	Directors of political, union, and civil organizations	1.1	2.4	2.26
	Officials, legislators and government officials	1.9	11.8	6.29
Professionals	General practitioners and specialists	9.7	9.7	1.01
	Fashion, industrial, and graphic designers, and interior decorators	3.2	3.2	1.02
	Other health specialists	2.9	3.2	1.11
	Specialists in agronomic sciences	3.6	4.3	1.19
	Broadcasters, entertainers, and clowns	1.6	1.9	1.22
	Professors, higher education instructors, and upper secondary teachers	10.2	13.9	1.37
	Other teachers and specialists in teaching, not elsewhere classified	2.4	4.8	2.02
	Performing artists	6.2	12.8	2.1
	Researchers and specialists in human sciences	2.8	5.8	2.12
	Educational supervisors and specialists in educational sciences	1	4.3	4.4
Technicians and associate professionals	Education aids and technicians, instructors and trainers	13.9	14.6	1.05
	Primay and secondary/middle school teachers	38.5	61.7	1.61
Clerks	Files workers and workers in control of stores and warehouses	27	30.7	1.14
	Supervisors of workers who provide and manage information	1.4	2.3	1.65
	Enumerators and encoders	2.4	8.3	3.53
Service workers and shop and market sales	Workers in the care of people	0.2	0.2	1.04
	Traders in stores	20	23	1.15
	Hairdressers, stylists, and related workers	1.1	1.3	1.22
	Gardeners	4.6	6.2	1.37
	Workers in the preparation and serving of food and drinks in establishments	12.9	17.6	1.38
Skilled agricultural and fishery workers	Workers in silvicultural and forestry activities	3.4	3.8	1.11
	Workers in agriculture	78.8	90.6	1.15
Crafts and related trades workers	Workers in the production and processing of food, beverages, and tobacco	6.8	8.1	1.21
	Artisans and workers in the manufacture of ceramics, glass, tile, and related	2.5	3.1	1.24
	Bricklayers and other workers in building construction	35.1	52.5	1.5
	Artisans and workers in the production of textile products	1.9	2.9	1.54
	Other craft workers, not elsewhere classified	0.9	1.4	1.58
	Artisans and workers in the production of wood products	5.6	9.7	1.73
	Weavers and workers in the preparation of textile fibers	0.3	1.3	5.2
Plant and machine operators and assemblers	Drivers of motorized land transport	57.2	63.6	1.12
	Operators of machinery for extraction in mines, quarries, and pits	2	2.3	1.16
	Operators of agricultural and forestry machinery	1.6	2.2	1.42
	Operators of machinery in the production of textiles, leather, and fur	5.7	9.5	1.69
Elementary occupations	Support workers in forestry, fishing, and hunting activities	20.3	23.7	1.17
	Assistant gardeners	0.2	0.3	1.25
	Support workers in forestry, fishing, and hunting activities	1	1.6	1.71
	Drivers of cycling transportation vehicles and animal-powered transports	0.6	1.3	2.17
	Support workers in agricultural activities	12.8	28.1	2.21
Armed forces	Workers in the army	98.1	99.6	1.02

[1] Coordinators and department heads in health, educational, and social services, and grading judges.

[2] Operators of machinery and equipment for extraction in mines, quarries, and pits.

[3] Operators of machinery and equipment in the production of textiles, leather, and fur.

.
.
.

Table 11: Summary Statistics

Variable	Obs	Mean	SD	Min	Max
Household variables:					
Children speak native	225,183	0.646	0.478	0	1
Only mother speaks native	225,183	0.093	0.290	0	1
Only father speaks native	225,183	0.120	0.325	0	1
# Other HH adults speak native	225,183	0.145	0.445	0	11
% municipality share HH lang.	225,183	0.521	0.314	0.0000015	.97
Mother's years of educ.	225,183	4.822	3.879	0	18
Father's years of educ.	225,183	5.468	3.917	0	18
Mother's age	225,183	38.751	8.009	25	54
Father's age	225,183	42.527	9.751	12	100
Normalised HH wealth	225,183	0.042	1.009	-1.481	3.483
Urban household	225,183	0.311	0.463	0	1
Language group variables (n=33):					
Employment return	225,183	0.028	0.033	-0.048	0.104
Group size (# of households)	225,183	236002	209959	2362	596636
Average wealth in group	225,183	0.314	0.439	-0.538	1.151
Average yrs. of education in group	225,183	4.960	0.659	3.131	7.261
Municipality variables (n=1962):					
Educational deprivation index	225,183	33.062	10.633	5.1	60.6
Health deprivation index	225,183	13.752	6.722	0.9	77.4
Housing deprivation index	225,183	32.188	16.935	1.3	82.7
Food deprivation index	225,183	27.960	12.311	0.5	85.7

Table 12: Language Transmission by Family Structure.

Language	Both parents Bilingual	Single Mother Bilingual	Mother Bilingual Father Spanish	Mother Spanish Father Bilingual	Parent-Child Pop
Nahuatl	.684	.512	.082	.048	367,488
Maya	.494	.31	.053	.025	129,570
Tzeltal	.937	.784	.161	.151	207,892
Tzotzil	.945	.835	.135	.11	175,270
Mixteco	.75	.624	.071	.045	122,274
Zapoteco	.688	.524	.074	.038	85,842
Chol	.923	.742	.113	.098	79,222
Totocana	.731	.54	.065	.037	57,926
Mazateco	.747	.466	.051	.057	55,072
Huasteco	.818	.602	.059	.071	46,506
Tlapaneco	.897	.763	.155	.099	44,256
Otomi	.427	.33	.043	.035	43,132
Purepecha	.735	.637	.169	.072	32,370
Chinanteco	.805	.534	.074	.06	31,234
Mixe	.797	.595	.095	.037	30,162
Tarahumara	.851	.566	.195	.05	18,384
Amuzgo	.921	.789	.11	.095	18,172
Zoque	.781	.566	.105	.04	17,880
Mazahua	.26	.213	.049	.026	17,206
Tepehuano	.96	.789	.214	.318	16,664
Huichol	.947	.783	.136	.064	16,474
Chatino	.901	.832	.075	.076	16,318
Tojolabal	.863	.804	.043	.07	15,954
Popoluca	.818	.637	.057	.053	13,210
Cora	.968	.875	.378	.092	8,838
Triqui	.811	.794	.072	.08	8,570
Huave	.871	.745	.074	.059	5,740
Chontal ⁛	.494	.4	.045	.069	5,054
Popoloca	.702	.529	.071	.008	4,444
Pame	.901	.875	.197	.05	3,666
Yaqui	.927	.652	.257	.064	3,478
Mayo	.284	.117	.107	.037	2,844
Cuicateco	.625	.358	.102	.026	2,502
All	.729	.553	.079	.047	3,052,206

Share of children from each parental linguistic arrangement that learnt native language of the parents.

⁛ Chontal de Tabasco. Data: Households (that identify as indigenous) and in which an indigenous language is spoken by either of the household parents. Mothers are 25-55 years, children under 5 excluded. Census 2015.

Parent-Child Pop: differs from number of observations because of weighting in sample.

.
.
.
This table shows a decomposition of households by years of schooling and language structure of the parents. The measure given is the transmission rate of the language within households. The rates (range: 0-1) of transmission decrease with the education of the parent.

Table 13: Language Transmission by Family Structure and Parental Education.

Years of schooling (Father)	Both parents bilingual	Mother bilingual father Spanish	Mother Spanish father bilingual	Parent-Child population
None*	0.845	0.154	0.065	403,942
1 year	0.754	0.107	0.049	81,154
2 years	0.779	0.085	0.049	182,264
3 years	0.755	0.092	0.058	260,452
4 years	0.773	0.074	0.054	139,114
5 years	0.754	0.082	0.046	114,034
6 years	0.741	0.085	0.045	703,016
7 years	0.689	0.09	0.046	22,326
8 years	0.647	0.052	0.049	43,380
9 years	0.601	0.056	0.045	478,978
10 years	0.567	0.031	0.057	12,644
11 years	0.519	0.042	0.057	14,204
12 years	0.486	0.075	0.039	125,030
13 years	0.64	0.044	0.024	5,598
14 years	0.349	0.044	0.015	4,272
15 years	0.441	0.059	0.044	11,724
16 years	0.441	0.08	0.063	4,2974
17 years	0.419	0.084	0.032	19,944
18 years or more	0.431	0.142	0.018	13,792

Share of children from each parental linguistic arrangement that learnt native language of the parents. Mothers are 25-55 years, children under 5 excluded. Census 2015. Parent-Child Pop: differs from number of observations because of weighting in sample. * Labelled as 'Non or preschool'.

.
.
.
.

Table 14: Language Transmission by Family Structure and Occupation of the Father.

Occupation of Father	Both parents Bilingual	Mother Bilingual Father Spanish	Mother Spanish Father Bilingual	Parent-Child Pop
Agriculture*	0.841	0.117	0.082	1127698
Construction	0.57	0.07	0.033	325154
Wholesale and retail	0.452	0.045	0.026	190340
Manufacturing	0.58	0.088	0.028	165222
Transportation$^{\diamond}$	0.49	0.062	0.046	74766
Hotels and restaurants	0.357	0.048	0.021	63160
Education	0.498	0.072	0.059	58378
Public administration *	0.447	0.068	0.023	52452
Other services	0.393	0.047	0.034	50852
Business services$^{\bullet}$	0.312	0.054	0.022	37992
Health and social work	0.369	0.064	0.018	8628
Private house services	0.406	0.067	0.032	8274
Mining and extraction	0.618	0.028	0.073	4216
Electricity, gas, water$^{\odot}$	0.527	0.054	0.006	4164
Financial services	0.223	0.022	0.01	1962

Share of children from each parental linguistic arrangement that learnt native language of the parents. * Agriculture, fishing, and forestry. $^{\diamond}$ Transportation, storage, and communications. * Public administration and defense. $^{\odot}$ Electricity, gas, water and waste management.$^{\bullet}$ Business services and real estate. Mothers are 25-55 years, children under 5 are excluded. Census 2015. Parent-Child Pop: differs from number of observations because of weighting in sample.

Table 15: Language Transmission by Family Structure and Occupation of the Father.

Language Group	Both parents Bilingual	Mother Bilingual Father Spanish	Mother Spanish Father Bilingual	Parent-Child Pop
Agricultural workers (industry code: 10)				
Nahuatl	0.767	0.105	0.07	153452
Tzeltal	0.97	0.225	0.221	146580
Tzotzil	0.962	0.191	0.137	102774
Chol	0.952	0.17	0.162	60512
Maya	0.684	0.092	0.04	59024
Mixteco	0.83	0.112	0.071	40474
Totocana	0.85	0.138	0.071	31018
Mazateco	0.903	0.141	0.148	30996
Zapoteco	0.775	0.09	0.055	30988
Tlapaneco	0.953	0.127	0.22	19664
Construction sector workers (industry code: 50)				
Nahuatl	0.581	0.078	0.037	35048
Maya	0.461	0.068	0.027	21068
Tzotzil	0.888	0.107	0.086	11108
Zapoteco	0.672	0.082	0.029	9482
Mixteco	0.57	0.064	0.045	7302
Tzeltal	0.691	0.061	0.108	5578
Otomi	0.375	0.037	0.024	5424
Totocana	0.517	0.042	0.031	4444
Mazahua	0.259	0.014	0.019	3296
Huasteco	0.733	0.061	0.105	3270
Crafts and related trade workers (occupation code: 7)				
Nahuatl	0.599	0.085	0.032	41282
Maya	0.395	0.044	0.023	20592
Zapoteco	0.619	0.061	0.027	9798
Tzotzil	0.912	0.087	0.034	8838
Purepecha	0.761	0.265	0.053	8474
Mixteco	0.544	0.05	0.032	6670
Otomi	0.383	0.028	0.029	6200
Tzeltal	0.68	0.029	0.105	5602
Huasteco	0.705	0.04	0.064	3914
Totocana	0.499	0.03	0.019	3908

Share of children from each parental linguistic arrangement that learnt native language of the parents. Mothers are 25-55 years, children under 5 are excluded. Census 2015. Parent-Child Pop: differs from number of observations because of weighting in sample.

Appendix 2: Institutional and Education Policies for Indigenous Populations in Mexico

Mexican society is a collection of groups with profoundly diverse backgrounds. One manifestation of this is the mosaic of languages that exist. While a significant number of languages have been lost through policies of cultural homogenization that began during colonial times and persisted into well the second half of the twentieth century, as of 2015, there were 66 Mexican languages written and spoken in different parts of the country (7.4 million people).

Indigenous language speaking populations tend to be concentrated in tight geographical areas, a pattern that appears to be fundamental for the transmission of language, one of the core points studied in this research.

From the end of the 19th century and much of the 20th century, educational policies for indigenous groups in Mexico were viewed as tools for crafting a homogeneous national identity around the idea of mestizaje.[15] Public institutions aimed for the cultural assimilation of these population groups through schooling and teaching of Spanish language. Indigenous cultures during the time were relegated and, common to the times, approached as if inferior to European ones (Stavenhagen, 1988, Salmerón and Porras 2010).

Organizations of native populations appeared after the revolution (1910-1917). Many of them originated in the 1930s but they only gained strength in the early 1970s. It was in this later decade in which public education shifted to adopt a multicultural and multilingual approach. Federal resources destined for the National Indigenous Institute grew more than tenfold between 1971 and 1976 (Sarmiento 1985) and in 1975 the first National Congress of Indigenous Populations was held.[16] The congress was the catalysing event for the creation of the National Council of Indigenous Populations[17] where, for the first time, representatives of indigenous groups would work together in a national political organization (Recondo 2007).

The creation of this national indigenous council, together with the debate about multicultur-

[15]The term "Mestizo" is a racial categorization from colonial times used to refer to a descent of a combined Spanish and American. This concept ignored the fact that within each region of the continent, now Latin America, independent cultures and civilization prevail.

[16]Occurred in Pátzcuaro, Michoacán.

[17]CNPI Spanish acronym.

alism and education of the time, led a series of institutional changes. In particular, bilingual education became a goal in itself rather than a vehicle for cultural homogeneity (García Segura 2004, Jiménez and Mendoza 2015, Jiménez-Naranjo and Mendoza-Zuany 2016).[18]

In 1978, a reform established that education would be imparted in the mother tongue of the child at least during the first years of primary school. The new focus on education would initially look only at the linguistic component as a differentiator, relegating the cultural element. Implementation of the reform took time due to technical difficulties but in 1984 textbooks, programs, guides, learning material and general books in over 20 indigenous languages were produced (Salmerón and Porras 2010). This material was created for preschooling and the first four years of primary school.

The next set of reforms occurred as a result of political pressure during the 1990s, a period that also saw the EZLN uprising (an ideological and armed movement led by indigenous in the state of Chiapas in 1994). Among the most significant accomplishments in favour of indigenous groups was a reform in Jan 1992 (Art 4)[19] recognising the constitutional right of indigenous communities to self-determination. The reform aimed to guarantee the right of these groups to preserve and enrich their languages, knowledge and culture. This reform would have important governance and administrative changes for indigenous communities long after.

In January 2001 the Federal Government created a national institute to coordinate bilingual and intercultural education (https://eib.sep.gob.mx/). This institute is in charge of developing educational curricula to attend cultural diversity, forming teachers, producing learning material, and pertinent school models. The Law of Linguistic Rights (2003), grants students of basic education the right to receive their education in their mother tongue, regardless of their location, a legal upgrade to the 1978 reform outlined above (Schmelkes 2006).

As can be seen, the institutional framework for the protection of indigenous languages is limited. Formal mechanisms for the protection of cultural identity need to be accompanied

[18] This shift was driven by organizations such as the National Alliance of Bilingual Indigenous Professionals which was founded in 1977.

[19] In current Mexican Constitution, as a result of another constitutional reform in 2001, the changes of Article 4 in 1992 have been shifted to belong to Article 2.

with resources for them to be effective. The policy mix of the kind seen in European nations is not really present in the context and, as a result, much of the transmission rely on informal mechanisms, mainly those of the household and the society in which individuals live.

The composition of the household and the characteristics of society are central in explaining the transmission of language. The role that each of these networks play is to some extent distinct. The easiness of learning a language will be a function of how the household is composed, that is of how many other indigenous language speakers there are in the household. As for societal networks, whether they are employment networks or of a more casual nature, these will be in the core of how valuable a language is. From an individual perspective, the value of a language is an increasing function of the number of other actual speakers. This is an example of spillover effects that is consistent with the existence of indigenous language clusters.