Heritage Language Labor Market Returns: The Importance of Speaker Density at the State Level

Christopher D. Blake* Daniel R. Walter†

October 21, 2019

Abstract

While certainly not a new position, vocal public opinion in the United States encourages immigrants to adopt English as their new standard of communication. Evidence for English-First (EF) movements is clearly visible on social and mainstream media (Volkova, Ranshous and Phillips, 2018). Even within immigrant communities, the support for heritage language use is an embattled topic (Seals, 2018; Liang, 2018). In this analysis, we ask whether there is an economic benefit, beyond and in addition to the social and cognitive benefits, of maintaining bilingual communal practices. Specifically, we are investigating if, and if so to what extent, continuing the use of a heritage language in a predominantly monolingual economy is monetarily beneficial. This paper adds a unique spatial dimension to the already vibrant literature on the returns to language by looking at the proportion of a region's population that natively (i.e. in the home) speaks a non-majority language. Using the Public Use Microdata Sets, we find evidence of a concave relationship between the proportion of a state's population speaking a heritage language and various economic outcomes, controlling for other factors. The results suggest a "sweet spot" population proportion in which benefits to speaking a heritage language are maximized. For a given region, our results have clear policy implications as states seek to target new in-migrant populations, as well as a more nuanced understanding of the relationship between minority-languages and their utility for their speakers. Of particular social importance are the insights these data provide in the backdrop of EF movements.

*Oxford College of Emory University, 801 Emory Street, Oxford, GA 30054, christopher.blake@emory.edu.
†Oxford College of Emory University, 801 Emory Street, Oxford, GA 30054, daniel.walter@emory.edu.

English is the most commonly spoken language in the United States, yet the nation possesses no federally official language (Citrin et al., 1990). This has not stopped multiple movements throughout the country's history from attempting to codify English as such (Ibid.). Because the federal government has avoided defining a national language, most efforts to affect change on this front come at the state level. These so-called "English-First" (EF) movements attempt to use policy to decree the language to be used in public administration activities such as education and regulation–largely avoiding discussions of the private sector for fear of over-infringing on personal liberty.[1] U.S. English, Inc. is currently the largest, non-partisan EF movement in the United States with over 1.8 million followers (U.S. English, Inc., 2019). Generally speaking, the political movement has focused on unification of the country as its primary driver, noting that assimilation into American culture is made more difficult for people that do not know the language.[2] Currently, such political efforts have resulted in 32 states with codes or constitutional amendments that outline English as the official regulatory language to varying degrees.[3]

While this call for a unifying language in the U.S. may appear well-intentioned, there are a number of social and political reasons against it which have been well articulated in Citrin et al. (1990), among others. Cost to heritage language speakers seems to be at the forefront of the debate surrounding speakers of other languages than English in a majority monolingual English-speaking environment like the United States. However, people looking for efficiency in government and the assimilation of immigrants and refugees through an English-first or English-only policy, as well as those looking to protect the rights of individuals (one among them being a right to your own language), have not fully investigated the positive possibilities afforded to speakers of minority languages in such an environment. By focusing on the positives, we believe that the narrative surrounding minority language use can shift to highlight the unique opportunities that minority and heritage languages provide their speakers.

To this point, we have used the terms "minority" and "heritage" as ways to describe non-English languages in the U.S. For our study, we will be using the term "heritage" to talk about the languages

[1] It should be noted that some have referred to these movements as "English-only" movements. These are not interchangeable as English-only movements advocate for laws that prohibit the use of languages other than English in both the public and private sphere. EF movements generally focus on making English the sole language used in the public sphere.

[2] It should also be noted that U.S. English, Inc. was founded by S.I Hayakawa, a Canadian born immigrant of Japanese ancestry who served a six year term in the Senate.

[3] See Appendix A for a state-by-state listing of when states implemented their EF laws.

used by speakers. We define a heritage language as the primary language an immigrant spoke in their country of origin prior to coming to the United States. Romas (1991) shows that immigrants in the U.S. tend to fully adopt English as their dominant language within three generations, however first and second generation U.S. immigrants tend to continue speaking their heritage language in tandem with English. Logically, it must be the case that individuals continue speaking these heritage languages because there are some returns associated with them–otherwise language assimilation would likely happen more quickly. This paper quantifies those returns for workers in the United States through their labor market outcomes.

We theorize these returns to be concave with respect to the proportion of heritage language speakers in the population. Specifically, if an individual speaks a language that is uncommon relative to the size of the population, then gains would be relatively low as there will be fewer opportunities for either network or human capital effects that could accrue to the worker. As the proportion of the population that speaks that particular heritage language rises, knowledge of the language itself becomes valuable through the network and market opportunities available in, and only in, that language which is accessible to, and only to, people with the necessary linguistic and cultural tools. Surrounded by a majority monolingual environment, speakers within a large enough community will, through their network, have access to the entire monolingual market, as well as a secondary, heritage-language specific market unavailable to the wider population, including both majority-language monolingual speakers and speakers of other heritage languages. In other words, the creation of an economically significant secondary market results from a large enough community of heritage language speakers who are the primary individuals capable of accessing that market.

Going along with our theory of a concave relationship between speaker density and returns to those speakers, we intentionally argue for a down-turn or plateauing of these returns as speaker density increases. We contend that this expectation of diminishing returns beyond a density "sweet spot" is likely because of the increasing availability of the market to a larger number of people within an area. That is, as long as the density of language speakers is not so high that the physically shared regional space eliminates any uniqueness, we expect there to be positive and increasing returns to heritage language speakers. There is also the possibility that the secondary market becomes accessible to the monolingual speaking community through extensive cultural and linguistic exposure after a certain density of heritage speakers inhabit a particular region,

allowing monolingual speakers the ability to navigate the secondary market to a sufficient degree that heritage-speakers are no longer advantaged. In either case, we expect that after a certain point, the returns for heritage speaker density will diminish.

Using the plethora of microeconomic data available through the Public Use Microdata Sets (PUMS), we find evidence to support this concave relationship, controlling for other factors. The results quantify these gains to heritage language speakers stemming from the density of their language. Evidence suggests the concave relationship exists for both employment income and a broader definition of income overall. Therefore, the paper represents an important contribution to labor, linguistic, and regional fields as we better understand the impact that language usage can have spatially in labor markets. From the workers' perspective, the results suggest an optimal regional population size that speaks a given heritage language, which has applications in local policy, as policymakers can better understand the potential effect that attracting immigrants may have on the constituents already in their community.

The remaining sections of this paper are as follows. Section 1 discusses the relevant literature and previous studies that analyze the labor market returns of languages. Section 2 outlines the econometric model we use to address the research question and briefly defines our expectations from the model. Section 3 describes the data applied to the econometric model and includes some brief summary statistics. Section 4 thoroughly discusses the results of the model and our evidence of the concave labor market returns to heritage languages. Finally, Section 5 concludes with some final thoughts about potential applications and avenues for future research.

1 Previous Work on the Relationship Between Language and Labor Market Returns

The connection between language ability and economic returns is one that is difficult to tease apart. Language is not only the way we communicate with our network, but the different languages we speak and how we speak them to whom are all indicators to our interlocutors during communication. Thus, a language does not simply indicate to someone which code we want to communicate in, but also reveals (or at least is interpreted as revealing) an abundance of information about ourselves, including age, gender, socioeconomic status, education level, emotional state, and more. All of

these perceived factors play a role in the way others interpret our language, and all of these factors are important indicators of economic returns. In this section, we review studies that have tried to isolate language use, or the use of a particular language in relation to others, and the ways in which language as a complex variable can affect an individual's returns.

In one attempt to corral the multiple co-variates of language and still glean insights into language's individual impact, Gonzalez (2005) looked at useful upper and lower bounds on limited English proficiency returns. For their data, Gonzalez (2005) looked at English proficiency as a self-reported measure of Hispanics living in the US. The author found that meaningful upper bounds on the effect of English proficiency on returns were education, occupation, and immigration status, while age of arrival was a useful lower band. A finer-grained reporting shows that returns on limited English proficiency are highly influenced by within-category differences. For example, differences in wage between well- and very well-spoken English is small at high education levels, but broader at other education levels, or differences between good and well spoken English within a particular occupation might be small, but large in another occupation. While this study provides some useful insights, it is limited in scope to Hispanic populations and focuses wholly on the use of the majority language by speakers of other languages, rather than their use of Spanish or other indigenous or heritage languages spoken by this population.

In another study looking at language use and its relationship to returns, Garrouste (2008) investigate a European context. The population of focus for this author is the returns for additional (i.e. second, third, and beyond) languages. Using three human capital specifications to evaluation returns and a Mincer approximation for experience, Garrouste (2008) find that for each of the eight countries compared, skills in a second language are estimated to be a major factor constraining wage opportunities. In the two countries in the study that have a national policy of bilingualism, specifically Finland and Norway, linguistic skills in the national, official language and in a second language have a significant effect on wages. In relation to our own study, the linguistic context of the speakers is quite different, but that is only the case if one perceives the United States as a fairly linguistically-homogeneous society. Broadly speaking, English is ubiquitous in the United States, but on the local level, the regional differences in languages is quite stark.

Turning specifically to the Nordic countries, Rooth and Saarela (2007) showed that prior knowledge of the country of immigration prior to arrival can be an important estimator of returns

to language. In their study, Rooth and Saarela (2007) investigated two groups of immigrants from Finland that had moved to Sweden. The first group had Finnish as a first language and the second group came from an area of Finland where a dialect of Swedish is used as the primary language of communication. There were significant differences in the returns to each group, which is to be expected. However, the researchers found that educational distribution between the groups was the largest indicator for the outcome differentials. There was also a difference within the Finish as a first language group by sex, where women had, on average, better second language Swedish abilities than the men. One important implication from their research on policy is their claim that, "Because decisions about how much governments should spend on language training for immigrants should be based on correct estimates for the returns to language skills, these issues are highly relevant also from a policy perspective," (Rooth and Saarela, 2007, pg. 209). In our paper we also intend to make policy recommendations, which is an important aspect to consider across linguistic, geographic, political, and cultural contexts. Another aspect that this paper highlights, which we previously mentioned, is the connection between language and demographic information. In this study, the authors specifically point to the Finnish immigrants with a Swedish dialect as their first language and discuss how, despite having the same first language, native Swedish citizens identify the Finnish immigrants as such and therefore belonging to a non-citizen group. This is an important touch-point that we will return to in our analysis of our findings.

A final paper by Bertrand, Luttmer and Mullainathan (2000) helps to frame our approach to the issue of returns to language. In this paper, the authors use language as the marker for culture in order to model the relationship between one's culture and the likelihood of that person ending up on welfare in the United States. While the use of language for a singular cultural marker is a bit tenuous, Bertrand et al.'s analysis is useful in our purview, in that language is clustered by location across the U.S. This analysis of language and economic outcomes through a spatial dimension has significant potential to inform our own methodology.

2 Methodological Approach to Heritage Language Effects on Labor Market Outcomes

With the previous literature in mind, our paper employs a unique approach and data set to ascertain the role of heritage languages in labor market outcomes such as employment and income, controlling for other factors. Consider two otherwise equal labor market participants with the same set of skills, education, experience, and so on; the only real difference between them being that one speaks English exclusively (English-only worker) and the other speaks a combination of English and a heritage language (heritage language worker). As seen in the previous literature described in Section 1, knowing a heritage language may imbue the speaker with both positive and negative effects. Let us then consider the how these effects may alter the labor market outcomes of the heritage language worker relative to that of the English-only worker. There may be positive network effects that allow the heritage language worker access to unique labor markets they find through those that speak the same heritage language. This would result in potentially higher pay, greater likelihood of being employed, as well as insulation from unemployment. Furthermore, employers would look favorably on an individual with additional language skills as the firm increases its interactions with consumer groups that speak a heritage language. From the worker's standpoint, then, the returns to their language ability will depend primarily on the proportion of heritage language speakers in their region.

The crux of our hypothesis is that language returns will be low both when the proportion of heritage language speakers to the population is low and high. In between these low and high proportions, however, a given worker will see increasing returns to their heritage language such that the relationship between returns and proportion takes on an "inverted U-shape." Figure 1 displays this paper's testable hypothesis. When the density is relatively low, there will be negligible benefits that accrue to the heritage language worker relative to an otherwise equal English-only worker. This stems from the fact that there will not be a sufficient network of heritage language speakers to expand the number of employment opportunities for the worker. In addition, employers are unlikely to value the language skills themselves if there are very few speakers in a region. This reduces the likelihood that a given worker would see a human capital premium from their language skills.

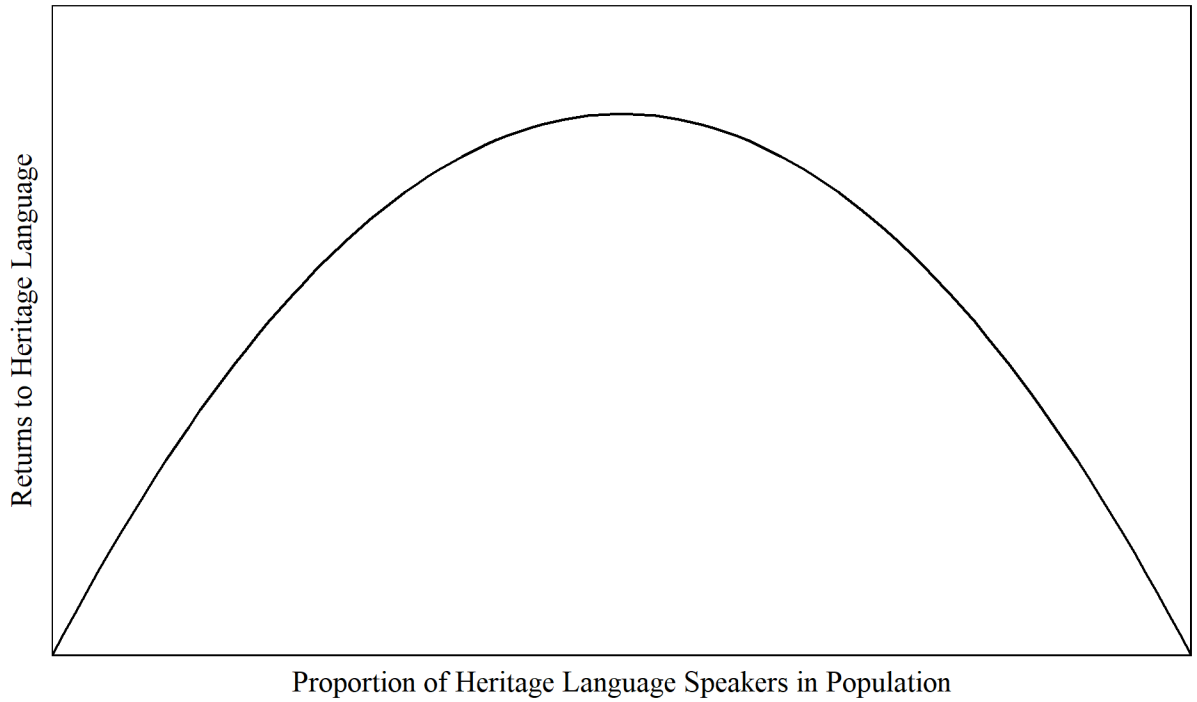

Figure 1: Relationship between Returns to Heritage Language Speakers Relative to English-only Speakers and the Proportion of the Population Speaking the Same Language

As the proportion of heritage language speakers in a region rises, the potential for positive network and human capital effects rises in turn. A worker is more likely to garner returns from specialized networks of those who speak the same language that provide employment opportunities. This effect would be particularly important for first-generation immigrants as they seek to find their footing in the United States. Secondly, as the proportion of heritage language speakers relative to the population rises, the language skills themselves become increasingly valuable for employers. Through their ability to communicate with a growing number of individuals in the region, these heritage language speakers can then provide a direct benefit to their firms. They would then likely see a human capital boost to their wage relative to an otherwise equal English-only worker.

We argue that this phenomena of a positive relationship between the proportion of speakers does not continue indefinitely, however, as the gains from the speaker's language dissipate due to two primary factors. First, there are likely diminishing returns associated with the potential benefits of speaking a given heritage language. As a higher proportion of the population speaks a language, it is less likely that the language is valued as highly in the labor market; there is a greater pool of workers with those language skills from which employers can choose. Second, as the proportion of

speakers grows for that particular language, the worker faces increased labor market competition for jobs that may require or recommend workers that have knowledge of the language. There are also likely diminishing returns to any network effects that one particular individual receives from membership to the heritage language network. Take, for example, a law firm in the Southwest of the United States. It is unlikely that the law firm will require more than one or two workers who speak Spanish fluently in order to adequately provide services to a significant portion of the population in those areas. A Spanish speaker would then not see returns associated with their language skills if the proportion of Spanish speakers to the population were exceedingly small (due to a dirth of need in hiring workers with that particular skill set) or large (which would increase the competition a given Spanish speaker would face for jobs that may find value in their language skills). While an obviously simplified example, it does highlight the point that a heritage language speaker is likely to face sufficiently high competition as density increases that will negate any potential benefits of their language skills.

The story may get more complicated if one considers potential negative effects of speaking a heritage language. It is easy to imagine the stigma and discrimination that certain Middle Eastern language speakers experienced in the wake of the September 11 attacks (for a recent example, see Hanassab, 2006; Akbaba, 2009). Our contention is that while this stigma may regrettably reduce the returns to speaking a given heritage language, the fundamental relationship between proportion of speakers and returns remains. For heritage language speakers that experience effects of stigma and discrimination, the returns function of Figure 1 would shift downward–indicating a reduction in the returns to language. It is likely still the case, however, that there is some form of the theoretical concave relationship as these groups still benefit from network and human capital effects (though perhaps to a lesser extent[4]).

2.1 Modeling the Theoretical Relationship Between Proportion of Heritage Language Speakers and Labor Market Returns by State

Our model is one of income and employment determination as a function of the proportion of the population speaking a heritage language. We analyze the years 2014-2016 to separate out any

[4]Although the argument could be made that they see even stronger effects for heritage language returns due to an extreme reliance on social networks

potential anomalies that may exist in a single year and to increase the sample size. Using the data described in Section 3, we consider the following general specification in Equation 1 for individual, i, in a given year, t, living in U.S. state, j:

$$Outcome_{ijt} = \beta_0 + \beta_1 P_{ijt} + \beta_2 P_{ijt}^2 + \delta W_{it} + \gamma D_{it} + \phi_t + \omega_j + \epsilon_{ijt} \quad (1)$$

P_{ijt} is defined as the proportion of the state population that speaks the same heritage language as the individual in a given year. To calculate this, we find the population of each state as well as the number of individuals speaking each language. P_{it} for an individual is then the number of people speaking that heritage language in the state divided by the state's population. W_{it} is a vector of individual work-related controls such as educational attainment, occupation, and experience (as estimated using the standard Mincer approach (Mincer, 1974; Psacharopoulos and Patrinos*, 2004), and D_{it} is a vector of individual demographic characteristics such as race, sex, migratory status, and English proficiency. Finally, ϕ_t controls for the year, ω_j is a fixed effect to control for the unobserved differences in wages across states, and ϵ_{ijt} reflects the error term. Outcome variables include wages and salaries and measures of weeks worked in the last 12 months.

The primary coefficients of interest are β_1 and β_2, which should have positive and negative signs, respectively, to support our hypothesis. To estimate these coefficients, we use Ordinary Least Squares and test for model specification using the Ramsey RESET test (Ramsey, 1969). Further tests for heteroskedasticity are performed to ensure the appropriate standard errors are employed (Breusch and Pagan, 1979). In order to ensure that the predicted signs of β_1 and β_2 are robust, multiple control specifications are used for the demographic and work-related variables used in our models.

3 Description of the Data Used for the Analysis

Our hypothesis is evaluated using the Public Use Microdata Sets (PUMS) available through the Census Bureau. This dataset contains demographic, economic, and location information about individuals in the United States and is intended to be representative of the full population. We use primary language spoken at home as a proxy for the heritage language of a worker in order to

Table 1: Summary Statistics–Labor Market Outcomes on Top, Determinants on Bottom

Variable	Observations	Mean	Std. Dev.
Wages	4,536,953	$46,873	$58,750
Weeks Worked Last 12 Months (1-6 scale, 1=whole year, 6=less than 14 weeks)	4,829,645	1.890	1.624
Language Proportion	11,528	0.184%	1.191%
Age (years)	9,342,552	41.485	23.354
Sex (2=female, 1=male)	9,436,102	1.511	0.500
Educational Attainment (years)	9,148,548	15.970	5.637
English Ability (1=high, 4=low)	1,639,036	1.651	0.926
Science or Engineering Field of Degree	2,154,805	1.638	0.481

assess the proportion of heritage language speakers in a given region. This provides nearly 3 million observations of individuals per year in the United States, over the years 2014-2016. Aside from income measures, the remaining variables in the dataset include educational attainment rates, race, gender, and field of degree as potential determinants of the income of an individual. Furthermore, there are variables corresponding to recent unemployment bouts, location, and occupation. The use of this data then allow for a sufficient number of controls to adequately assess whether the returns to heritage language workers are concave as predicted.

Over the years 2014-2016, Table 1 shows the number of observations, the mean values, and standard deviations of variables in the sample. The primary variable of interest is the Language Proportion variable, which takes on a mean of 0.184% across all heritage languages in all states. Unsurprisingly, Spanish is the most commonly spoken heritage language for each year in the sample. In New Mexico in 2014, approximately 27.7% of individuals primarily spoke Spanish at home, representing the highest value for Language Proportion in our dataset. By 2016, more than 20% of the populations in Texas, California, New Mexico, Nevada, and Florida spoke Spanish as the primary language used in the home. The least common languages vary by year and state, but tend to include Native American languages, some Oceanic languages, and African languages.

Because our analysis looks at each language spoken in every state, visualizing the density of heritage languages is difficult. There are 134 language categories for which the Census Bureau collects data and this implies 6,700 observations of the Language Proportion variable per year if we assume that every state has one person speaking each heritage language. There are a number of languages that regrettably do not have state-level observations in our dataset, so this does restrict

the size of our sample and motivates our decision to use multiple years worth of data rather than a single year. Depending on model specification, the limitation of the observed data reduces the number of observations from nearly 10 million, to between 200,000 and 800,000. Due to languages in the dataset with a limited number of observations, the distribution is skewed to the right with a cluster of languages at the state-level with small Language Proportions. We considered the effect of these high and low outliers and ultimately decided that there would be no justification for dropping either end of the Language Proportion distribution as such a change would be entirely arbitrary. Furthermore, to drop a heritage language from the observations (e.g. Spanish) would actually not allow us to speak to the downward sloping portion of our theorized relationship between Language Proportion and economic returns.[5]

4 Heritage Language Returns Results

Even with the relatively small variation in the Language Proportion variable, there is significant evidence of the concave relationship between Language Proportion and economic outcomes. Of note, our results show exceptional robustness, even with the existence of clear outliers (e.g. Spanish), adjusted standard errors to account for heteroskedasticity, and a wide array of model specifications. In the following sections, we outline the primary modeling results and our evidence for the predicted concave relationship.

4.1 Results with Respect to Wages and Salaries

Table 2 shows the relationship between Language Proportion and wages and salaries for workers in the sample. According to the fourth and most comprehensive regression, this would imply that a marginal increase in the Language Proportion of a worker by 0.01 (one percentage point) would change wages and salaries by approximately $(0.630 - 4.656 * LanguageProportion) * 100\%$. The un-weighted mean of this scaled Language Proportion in the sample is 0.118. From the mean, the model implies that a percentage point increase in the Language Proportion would generate an average return to wages of 8.01%, ceteris paribus. This is a significant increase in wages and

[5]Preliminary results indicate that if Spanish is removed from the sample, the statistical significance of a Squared Language Proportion is nonexistent. The remaining linear term has a statistically significant positive coefficient in these regressions, indicating we capture the initial network and human capital returns quite well and only through high proportion languages can we see the downward sloping portion of the inverted U-shape.

salaries, surpassed only by the effects of educational attainment and English proficiency in the models. The signs and statistical significance are consistent across each specification in Table 2 and reinforce previous wage determination model results. Specifically, each year of educational attainment increases expected wages and salaries while experience has an initially positive effect with diminishing returns for each additional year.

Table 2: OLS Regression Results–Wages and Salaries with Respect to Language Proportion

	Dependent variable:			
	Log of Wages and Salaries			
	(1)	(2)	(3)	(4)
Language Proportion	0.457***	0.546***	0.604***	0.630***
	(0.056)	(0.056)	(0.056)	(0.098)
Squared Language Proportion	−1.659***	−1.956***	−2.236***	−2.328***
	(0.216)	(0.215)	(0.215)	(0.381)
Educational Attainment	0.031***	0.030***	0.027***	0.125***
	(0.0003)	(0.0003)	(0.0003)	(0.002)
Mincer Experience	0.061***	0.061***	0.062***	0.061***
	(0.0002)	(0.0002)	(0.0002)	(0.0004)
Squared Mincer Experience	−0.001***	−0.001***	−0.001***	−0.001***
	(0.00000)	(0.00000)	(0.00000)	(0.00001)
English Ability			−0.057***	−0.162***
(1=Proficient, 6=Limited)			(0.001)	(0.003)
Science/Engineering Degree				−0.024***
(0=Yes, 1=No)				(0.004)
Year Controls	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Demographic Controls		Yes	Yes	Yes
Occupational Controls	Yes	Yes	Yes	Yes
Migration Controls			Yes	Yes
Observations	838,951	838,951	838,951	262,405
R^2	0.390	0.397	0.399	0.369
Adjusted R^2	0.389	0.397	0.399	0.368

Notes: Table made with Hlavac (2015) stargazer package in R; Significance denoted: *p<0.1; **p<0.05; ***p<0.01

The coefficients on Language Proportion are promising, but the natural question is what these coefficients imply about the upper echelon of returns to heritage language speakers. Specifically, what Language Proportion would maximize the returns to speakers of that language? The results of specifications 1-4 imply that Language Proportion would have a positive effect on wages until Language Proportion was 0.138, 0.140, 0.135, and 0.135 respectively. As noted previously, the unweighted average Language Proportion across states is 0.118 which implies that most heritage language speakers could benefit from an increase in the proportion of the population that speaks their language. Furthermore, the median of Language Proportion estimates is 0.07 (right skewed distribution) so more than 50% of heritage language speakers would benefit from an increase in the number of workers speaking the same language.

English ability is controlled for in the more robust specifications in order to prevent the Language Proportion variable from simply capturing the English ability of workers. Our hypothesis relies on the fact that otherwise equal workers who speak heritage languages have access to greater employment opportunities through networks not accessible to English-only workers. If English ability were not controlled for, Language Proportion may pick up on a given speaker's English ability rather than the prevalence of their language in the population. Because we are specifically trying to capture returns to heritage languages and not English proficiency, models three and four in Table 2 are likely to best capture the relationship we seek. Moreover, the negative coefficient on English ability is expected and supports the conclusions of Romas (1991), which provide evidence that most immigrants tend to adopt English in the United States over the course of three generations. The variable takes on a value of 1 for those proficient in English and 6 for those with limited English ability and so the negative coefficient implies that wages fall drastically for each reduction in English-speaking ability. The economic significance of the coefficient (16% wage and salary reduction for each categorical reduction in English proficiency is also likely the significant motivator for those who immigrate speaking only a heritage language to adopt English as quickly as possible.[6]

Turning to overall model specification, Ramsey RESET tests do indicate a p-value well below

[6]The English proficiency variable in the PUMS dataset is self-reported and so it should be noted that this may partially drive the results. For example, an over- or understatement by any particular language speaker group could bias results. Studies in second language self-assessment indicate that its reliability relies on a number of factors, including personality and psychological traits (AlFallay, 2004), experience in self-assessment and monitoring (Ross, 1998; Blanche and Merino, 1989), and anxiety in using the language (MacIntyre, Noels and Clément, 1997). This indicates there is likely a large amount of variability within the English proficiency variable.

0.01 for the wage and salary models. Thus we reject the null hypothesis of a correct functional form specification. However, we have controlled for region, time, experience, English ability, gender, education. To our knowledge, there is no theoretical justification for including cubed terms for experience or squared terms for educational attainment, for example. In addition, the RESET test is likely influenced by the number of controls in the models. We have 50 state controls, 3 year controls, and over 200 occupational controls. Combined, the RESET test is unlikely to accurately suggest the model is misspecified because adding squared and cubed versions of those variables would undoubtedly increase the predictive ability of the model. Therefore, while a common test may indicate misspecification, we focus on the relative proximity of the R-squared and adjusted R-squared terms for each model. Because the incorporation of more variables does not diminish the adjusted R-squared much, if at all for most models, this would indicate that the variables included are actually worthy of inclusion.

4.2 Results with Respect to Weeks Worked per Year

There is strong evidence of concave returns to heritage language speakers as the Language Proportion changes when looking at the wages and salaries of workers. Part of our contention is that workers with a specific heritage language are also privy to employment opportunities stemming from the network of heritage language speakers in their state that English-only workers and speakers of other heritage languages are not. We use number of weeks working in the past year as a proxy for employment opportunities. The hypothesis is that an increase in Language Proportion would initially increase the likelihood a worker is employed most of the year, but these effects would diminish as competition for jobs increased and the relative scarcity of heritage language speakers diminished. Table 1 shows that the Weeks Worked variable in the PUMS dataset is counter-intuitive, with lower values indicating that a worker was employed all 52 weeks and the highest value indicating a worker working 14 weeks or less. We adapt the categories to take on a value of 0, 1, or 2 prior to using them in the regressions with 0 meaning the worker worked less than or equal to 26 weeks, 1 meaning the worker worked between 26 and 47 weeks, and 2 meaning the worker worked 47 to 52 weeks.

Table 3 shows the OLS results for Language Proportion on the categorical weeks worked per year for each worker. Of note, the predictive ability of the model relative to those for wages and salaries

Table 3: OLS Regression Results–Wages and Salaries with Respect to Language Proportion

	Dependent variable:			
	Weeks Worked Per Year (0-2 Scale)			
	(1)	(2)	(3)	(4)
Language Proportion	0.335***	0.345***	0.355***	0.305***
	(0.020)	(0.020)	(0.020)	(0.035)
Squared Language Proportion	−0.700***	−0.726***	−0.783***	−0.722***
	(0.078)	(0.078)	(0.078)	(0.135)
Education Attainment	0.0003**	0.0002**	−0.0003**	0.006***
	(0.0001)	(0.0001)	(0.0001)	(0.001)
Mincer Experience	0.007***	0.007***	0.006***	0.007***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Squared Mincer Experience	−0.0001***	−0.0001***	−0.0001***	−0.0002***
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
English Proficiency			−0.008***	−0.014***
			(0.001)	(0.001)
Science/Engineering Degree				0.001
				(0.001)
Year Controls	Yes	Yes	Yes	Yes
State Controls	Yes	Yes	Yes	Yes
Demographic Controls		Yes	Yes	Yes
Occupational Controls	Yes	Yes	Yes	Yes
Migration Controls			Yes	Yes
Observations	791,660	791,660	791,660	253,278
R^2	0.047	0.048	0.050	0.058
Adjusted R^2	0.046	0.047	0.049	0.056

Notes: Table made with Hlavac (2015) stargazer package in R; Significance denoted: *p<0.1; **p<0.05; ***p<0.01

is far less. This is unsurprising given the categorical nature of the dependent variable rather than the nearly normal distribution of adjusted wages in Table 2. Despite this, the adjusted R-squared is close enough to the R-squared in each specification to suggest that each variable included should remain as the penalty for including regressors is sufficiently low. Furthermore, the signs of each coefficient remain consistent with the analysis of wage and salary determinants.

Turning to our primary hypothesis, Table 3 strongly suggests a concave relationship between Language Proportion and weeks worked per year. Because the dependent variable is not numerical, there is no value added in directly interpreting the marginal effect of an increase in Language Proportion. That being said, the regression results indicate that the likelihood of a heritage language speaker working full-time in a given year is maximized when the Language Proportion is between 0.21 and 0.23. Interestingly, there are significant implications with these results. First, we noted that more than 20% of workers in five states speak Spanish at home. For New Mexico in 2016, this percentage was 27%. At proportions so high, the returns to heritage languages begin to diminish for labor market participants and they would generally be better off with smaller Language Proportions.

The second implication stems from the fact that most languages have Language Proportions far lower than 20% in each state. Considering the nature of the sample, this means the diminishing effect of Language Proportion on weeks worked annually is experienced by a minority of heritage language speakers. In other words, while returns to the Language Proportion are statistically concave, they are not economically concave for the majority of heritage language workers. Most heritage language workers would benefit from an increase to the Language Proportion as it both increases the chance that they are able to work full-time and insulate themselves from bouts of unemployment (which would reduce the number of weeks worked).

The primary conclusion across these sets of models is that there are clearly diminishing returns for some heritage languages for wages and salaries. Due to the U.S. sample, nearly all languages, except for Spanish, benefit from language proportion with respect to employment. Because there is evidence of a concave relationship, it is important to know when the concave occurs to understand the maximum beneficial heritage speaker density. This finding not only supports our hypothesis that language returns diminish beyond a specific "sweet spot", but also has significant policy implications at the state-level as policymakers can better understand the potential costs and benefits of any action that attracts or repels heritage language speakers of given groups.

4.3 Limitations, Robustness Checks, and Suggestions for Future Research

Our results indicate promising evidence of a concave relationship between the proportion of heritage language speakers and economic outcomes. In critiquing our results to make sure they capture what

we aim to, this section considers potential confounding factors that may muddle our results. First and foremost, many of the languages in the PUMS dataset have a proportion of speakers less than 1% of the population in a particular state. To be more specific, of the 242 language categories in the PUMS dataset, 165 have language proportions less than 0.5% of the population. This results in a significant cluster of languages at the lower end of the language proportion distribution that may drive the results. Furthermore, Spanish is the only language that has language proportions above the regression-implied turning point of approximately 0.13–or approximately 13% of a state's population. The most obvious critique to our results would be that outliers then drive the concave shape we test.

The issue with simply dropping Spanish from the dataset is that we *expect* languages with a higher language proportion to receive limited economic benefits relative to native English speakers. To omit them would then be to lose a key portion of the story. As one solution, we consider quantile regressions. These regressions divide the Language Proportion variable into three, five, or ten categories, and allow us to eliminate the need to impose a nonlinear squared term in our functional form and let the data speak directly to the relationship between the proportion of heritage language speakers and economic outcomes. In order to support our hypothesis, we would expect that at lower quantiles, there exists a positive relationship between Language Proportion and our variable of interest. This effect should taper off and become statistically insignificant in later quantiles (indicating a null relationship and representing the top of the concave relationship in Figure 1) and finally become negative by the last quantile to represent the notion of diminishing labor market returns. The potential downside to quantile regressions as a solution is that they lump together multiple languages into blocks that may be weighted more heavily towards one language as opposed to another. As a result, the signs of our coefficients may not be universally expected.

Tables 4 and 5 show the results for analogous five-category quantile regressions with the full set of controls listed in the fourth equations of Tables 2 and 3 (i.e. educational attainment, potential experience, English proficiency, occupational, state, and year controls).[7][8] Table 4 exhibits an

[7]Five quantiles seems to best match the clusters of data and so we display those results here. The results were similar for the use of three quantiles, but the use of ten quantiles generated more inconsistent results with the higher number of categories.

[8]The observations in these tables are not uniform across quantiles because quantiles were generated using the Language Proportion variable and it is the case that some observations are subsequently dropped during the statistical analysis as a result of some missing control variables.

Table 4: Quantile OLS Regression Results–Wages and Salaries with Respect to Language Proportion

	Dependent variable:				
	Log of Wages and Salaries				
	Quantile 1	Quantile 2	Quantile 3	Quantile 4	Quantile 5
Language Proportion	-0.067^{***}	0.020^{***}	0.002	-0.023	-0.287^{***}
	(0.026)	(0.006)	(0.002)	(0.019)	(0.074)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	176,057	167,562	173,276	162,303	159,753
R^2	0.413	0.396	0.321	0.300	0.308
Adjusted R^2	0.411	0.394	0.319	0.298	0.306

Notes: Table made with Hlavac (2015) stargazer package in R; Scaled version of Language Proportion used such that one unit is equal to 1%; Significance denoted: *p<0.1; **p<0.05; ***p<0.01

initially positive relationship between Language Proportion and wages and salaries for heritage language speakers. For Quantile 2, a one percent increase in the Language Proportion is associated with a 2.02% controlled increase in wages and salaries while a similar increase in Quantile 5 implies a decrease of 24.95%–an economically significant amount. Such a steep decline in predicted wages and salaries over the fifth quantile demonstrates how quickly the benefits of heritage language networks diminish in states with large proportions of speakers. The effect of changing Language Proportion in Quantiles 3 and 4 is negligible, as expected, which leaves Quantile 1 as the only anomaly in the results displayed in Table 4. Indeed, the effect of increases to Language Proportion appear to hurt wages and salaries for languages with very few speakers.

The negative coefficient for Quantile 1 does not directly align with our theory, but suggests that those living in a monolingual culture like the United States may be worse off than otherwise equal peers until a critical mass of network size is available in their area. Indeed, the cutoff for a language to fall into Quantile 1 is 0.33% of the population and it is highly unlikely that heritage language speakers falling into this category are able to garner any sort of network benefits as described in our theory. As a result, we would expect a statistically null coefficient for Quantile 1 language speakers. One possible explanation for the negative coefficient could then be some sort of discrimination; however, this would be difficult to truly prove given the nature of our dataset. Another possibility is that prospective employers may view workers with language skills in a relatively unknown group

Table 5: Quantile OLS Regression Results–Weeks Worked per Year with Respect to Language Proportion

	Dependent variable:				
	Weeks Worked Per Year (0-2 Scale)				
	Quantile 1	Quantile 2	Quantile 3	Quantile 4	Quantile 5
Language Proportion	-0.414^{***}	0.276^{***}	0.040^{***}	-0.273^{**}	-0.336
	(0.156)	(0.034)	(0.011)	(0.126)	(0.496)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	187,789	179,764	182,846	174,483	172,997
R^2	0.120	0.111	0.100	0.074	0.079
Adjusted R^2	0.117	0.108	0.097	0.072	0.076

Notes: Table made with Hlavac (2015) stargazer package in R; Scaled version of Language Proportion used such that one unit is equal to 1%; Significance denoted: *p<0.1; **p<0.05; ***p<0.01

broadly as immigrants and therefore pay them less as a sort of risk "premium" that compensates employers for the potential costs of training and maintaining the worker. Given that we focus on network benefits, our theory cannot account for this negative coefficient and informs both a limitation of the present work, as well as the potential for future studies. It is not unreasonable to assume that heritage language speakers that represent exceptionally small proportions of the overall population face entirely different labor market challenges and benefits than those for whom the network has reached a critical mass.

With qualification, our theory holds with respect to wages and salaries for all but the first quantile of heritage language speakers. Our robustness checks for the number of weeks worked per year yield very similar results and are displayed in Table 5. Quantiles 2 and 3 provide evidence of the initially positive effect of increased Language Proportion on wages, while Quantiles 4 and 5 demonstrate the predicted negative effect. With reference to Table 4, we see a significantly negative effect on employment in lower quantiles than we do for wages and salaries. This suggests that the diminishing returns to increased heritage language networks occur more quickly while wages and salaries would continue to rise as employers increasingly valued the worker's language skill. Quantile 1 again shows a statistically negative effect of increased Language Proportion on economic outcomes. Here, the implication for a critical mass of heritage language speakers becomes more clear. In order to help these heritage language speakers accrue additional labor market benefits, policymakers

should foster and support an environment in which speaker networks are encouraged to grow.

As a further robustness check, we also tested for the sensitivity of the results to outliers. Because we did not want to drop a relatively common language such as Spanish, we focused on outliers of languages with very low Language Proportions. With a cutoff proportion of 0.005 (or 0.5% of the population) we performed the regressions in Tables 2 and 3 and got statistically significant and similar results as before. We also performed separate robustness check regressions on stratified samples of both exclusively Spanish speakers and Spanish heritage language speakers. For each of these sample variants, the predicted concave relationship results were identical to those above–adding further evidence for our theory. Specifically, whether the sample was stratified into Spanish speakers only, non-Spanish speakers, or a sample omitting speakers of uncommon heritage languages (below 0.5% of the population) the concave relationship between heritage language proportion and economic outcomes remains.

We hope that this novel approach to testing the relationship between language speaker density and economic returns on a regional scale can foster additional lines of inquiry. For instance, the largest limitation to our working theory about the relationship between language speaker density and economic returns is that we have only investigated one country. We hope that datasets from additional countries with different linguistic contexts and populations can contribute to our initial hypothesis and tease apart some of the tangled links between language speaker proportions, population sizes, and additional language acceptance and education. For example, Portuguese is one of the most widely spoken heritage languages in Japan, and Japan has a similarly monolingual-dominant population (Takenoshita et al., 2014). Do Brazilian immigrants in Japan face the same concave relationship as Spanish-speakers in the United States, or does the size of the overall population or the prevalence of another majority-minority language (in this case Chinese) play a factor in mitigating this downturn?

A second area of future research, to which we already alluded, is the economic outcome for small populations of heritage language speakers. In our case, we found a negative relationship between speaker density and economic outcomes in Quantile 1. Can this result be replicated in other countries, and if so, what exactly is driving this initial downturn? A mixed-methods approach would be required to shed light on this finding.

5 Brief Conclusions

The overall findings of our paper confirm our initial hypothesis about the relationship between heritage speaker population density by region and wage returns. We believe that our theory about secondary heritage-language market access provides the most plausible explanation for this relationship, given the importance of density by region over the effect of any particular language on its own. Whether looking at wages and salaries or likelihood of employment, there are initially positive returns to increasing the density of heritage language speakers that eventually dissipate as the labor market becomes saturated. This explains the phenomenon we observed in our introduction. Namely, there are returns to continuing to speak a heritage language and this represents a viable explanation for why workers do not immediately integrate linguistically into the primarily English labor markets in the United States. Our results demonstrate that these workers experience statistically significant wage increases in tandem with a higher chance of working full-time as the size of the heritage language population in their region rises. Such returns are concave, however, which implies a certain optimum Language Proportion.

The potential impacts of these findings are wide-reaching in terms of both policy decisions and general societal acceptance of heritage languages in minority-language speaking populations. While we do not claim that this economic benefit should be viewed as more convincing to a wider range of people than other arguments, we believe it weaves itself into a positive narrative on heritage language value, in the humanistic sense.

For policy makers, there is an important distinction to be made between the wage and employment outcomes for their constituents. While the mean for wages is near the concave and slightly right-skewed for language density, the mean for employment shows a much stronger right-skewed distribution, indicating significant gains for employment even for very small increases in language density. Employment has a significant impact on crime, entitlement programs, happiness, and the burden on systems outside of entitlement. If policy makers are interested in improving these areas, supporting minority individuals and communities that are affected to a greater degree by these issues will have a ripple effect that betters the life of every resident through stronger employment numbers.

In terms of direct policy implications, we hope that this paper can provide some insight into

the types of language speakers that communities and cities should be attempting to attract to their area. In reality, the networks of heritage language speakers are truly founded at the grass-roots level. Speakers find each other in order to access opportunities to use skill sets that they may not be able to make use of in an English-speaking environment (e.g. sales and service positions that require both professional and linguistic resources). In seeking out these opportunities with speakers of the same language, they develop a network for others also looking for employment opportunities. At the policy level, the goal should not be the establishment, which would likely be viewed as inauthentic, but promotion and advertisement of networks of heritage language speakers. This type of policy is crucial because it does not subvert the authenticity of the network, but rather acknowledges its importance as such without appropriation. Providing knowledge to heritage language speakers from other areas about potential occupational opportunities not available to them in their current location should drive inter-regional migration. At the maximum returns density, advertisement and promotion could be limited, but the market would likely show this reduction in opportunity on its own, which would reduce the influx of a particular group of heritage language speakers.

We also believe that this paper has the potential to influence policy makers and the general public to view heritage language speaking communities as socially, culturally, and economically vital to the fabric of the U.S. This view is much more in line with a historically accurate view of the U.S. in a time when many people interpret EF and English-Only movements as logical truths resulting from the myth of a monolingual America. In doing so, we hope to contribute, in some small way, to the deconstruction of the "In America we speak English" rallying cry and symbol of national unity in favor of a more historically accurate representation of the U.S. as a linguistically and culturally diverse nation.

References

Akbaba, Yasemin. 2009. "Who discriminates more? Comparing religious discrimination in Western democracies, Asia and the Middle East." *Civil Wars*, 11(3): 321–358.

AlFallay, Ibrahim. 2004. "The role of some selected psychological and personality traits of the rater in the accuracy of self-and peer-assessment." *System*, 32(3): 407–425.

Bertrand, Marianne, Erzo FP Luttmer, and Sendhil Mullainathan. 2000. "Network effects and welfare cultures." *The Quarterly Journal of Economics*, 115(3): 1019–1055.

Blanche, Patrick, and Barbara J Merino. 1989. "Self-assessment of foreign-language skills: Implications for teachers and researchers." *Language learning*, 39(3): 313–338.

Breusch, Trevor S, and Adrian R Pagan. 1979. "A simple test for heteroscedasticity and random coefficient variation." *Econometrica: Journal of the Econometric Society*, 1287–1294.

Citrin, Jack, Beth Reingold, Evelyn Walters, and Donald P Green. 1990. "The "official English" movement and the symbolic politics of language in the United States." *Western Political Quarterly*, 43(3): 535–559.

Garrouste, Christelle. 2008. "Language skills and economic returns." *Policy Futures in Education*, 6(2): 187–202.

Gonzalez, Libertad. 2005. "Nonparametric bounds on the returns to language skills." *Journal of Applied Econometrics*, 20(6): 771–795.

Hanassab, Shideh. 2006. "Diversity, international students, and perceived discrimination: Implications for educators and counselors." *Journal of Studies in International Education*, 10(2): 157–172.

Hlavac, Marek. 2015. "stargazer: well-formatted regression and summary statistics tables. R package version 5.2."

Liang, Feng. 2018. "Parental Perceptions toward and Practices of Heritage Language Maintenance: Focusing on the United States and Canada." *Online Submission*, 12(2): 65–86.

MacIntyre, Peter D, Kimberly A Noels, and Richard Clément. 1997. "Biases in self-ratings of second language proficiency: The role of language anxiety." *Language learning*, 47(2): 265–287.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2.* ERIC.

Psacharopoulos, George, and Harry Anthony Patrinos*. 2004. "Returns to investment in education: a further update." *Education economics*, 12(2): 111–134.

Ramsey, James Bernard. 1969. "Tests for specification errors in classical linear least-squares regression analysis." *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(2): 350–371.

Romas, Leo Johnathan. 1991. "English First legislation: Potential national origin discrimination." *Chicano-Latino L. Rev.*, 11: 77.

Rooth, Dan-Olof, and Jan Saarela. 2007. "Native language and immigrant labour market outcomes: An alternative approach to measuring the returns for language skills." *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale*, 8(2): 207–221.

Ross, Steven. 1998. "Self-assessment in second language testing: A meta-analysis and analysis of experiential factors." *Language testing*, 15(1): 1–20.

Seals, Corinne A. 2018. "Positive and negative identity practices in heritage language education." *International Journal of Multilingualism*, 15(4): 329–348.

Takenoshita, Hirohisa, Yoshimi Chitose, Shigehiro Ikegami, and Eunice Akemi Ishikawa. 2014. "Segmented assimilation, transnationalism, and educational attainment of Brazilian migrant children in Japan." *International Migration*, 52(2): 84–99.

U.S. English, Inc. 2019. *https://www.usenglish.org/*, Accessed: 2019-03-26.

Volkova, Svitlana, Stephen Ranshous, and Lawrence Phillips. 2018. "Predicting Foreign Language Usage from English-Only Social Media Posts." Vol. 2, 608–614.

A Appendix–EF Outline by State

The following is a summarized version of EF-codification as provided by U.S English, Inc. Entries represent the year in which most or all of official state business was to exclusively use the English language.

State	Year	State	Year
Alabama	1990	Montana	1995
Alaska	1998	Nebraska	1920
Arizona	2006	Nevada	
Arkansas	1987	New Hampshire	1995
California	1986	New Jersey	
Colorado	1988	New Mexico	
Connecticut		New York	
Delaware		North Carolina	1987
Florida	1988	North Dakota	1987
Georgia	1996	Ohio	
Hawaii	1978	Oklahoma	2010
Idaho	2007	Oregon	
Illinois	1969	Pennsylvania	
Indiana	1984	Rhode Island	
Iowa	2002	South Carolina	1987
Kansas	2007	South Dakota	1995
Kentucky	1984	Tennessee	1984
Louisiana	1811/1992	Texas	
Maine		Utah	2000
Maryland		Vermont	
Massachusetts	1975	Virginia	1996
Michigan		Washington	
Minnesota		West Virginia	2016
Mississippi	1987	Wisconsin	
Missouri	2008	Wyoming	1996

Table 6: Source: U.S. English, Inc. (2019)