Which American Languages are Dying? Quantifying the Demographic Vulnerability of Indigenous Languages in the United States.

INTRODUCTION

Language is a salient marker of sociocultural identity, and as such the historical trajectory of a particular spoken language is a vital part of understanding the future of its respective language speaking community, whether it is an ethnicity, tribe, or nation. Demographic concepts such as aging and population replacement are often invoked in the study of language vitality (the extent to which a language is positioned to continue being adopted and used in the future). For example, cohort replacement mechanisms play out as the new dominant tongue is taught to children in school systems and at home while speakers of the declining language age and eventually die (perhaps most prominently discussed in Abrams and Strogatz 2003; see also Rehg and Campbell 2018 and the UNESCO report on the issue [2003]). While some research has preliminarily touched on the theoretical role of demography in the study of linguistic vitality (for example, see Siegel 2018), very few applied attempts have been made to utilize demographic methods to inform our picture of the long-run vitality of different languages.

This paper will address the potential of performing cohort-component population projections of language speakers, as well as the use of other demographic indicators such as net reproductive rate, population momentum, and old-age dependency to inform our understanding of linguistic vitality. Specifically, I use the American Community Survey (hereafter ACS) to calculate these measures for a group of indigenous languages and discuss their respective benefits, drawbacks, and complications. This knowledge in turn could help quantitatively hone existing measures of linguistic vitality, and would ultimately help predict the futures of language speaking communities in much the same way that those measures help predict future ethnic and national populations. The indigenous American languages analyzed here are: Inupiaq, Central

Yupik, Western Apache, Navajo, Lakota, Eastern Keres, and Cherokee. Inupiaq and Yupik are native Alaskan languages. These represent the largest Native American language communities in the United States (according to the ACS). Besides Navajo, Yupik has the largest community of indigenous speakers in the United States today. Western Apache is spoken in Arizona. Navajo is spoken in both Arizona and New Mexico, and is the largest indigenous American language community in the United States by far. Lakota is based in the Dakotas, Eastern Keres is spoken predominantly in New Mexico, and Cherokee speakers are based out of the Cherokee tribal populations in North Carolina, Oklahoma, and other areas.

DEMOGRAPHIC PROCESSES

Changes in the size of language speaking communities can be conceptualized through a demographic lens. The death of a small language or the gradual displacing of a minority language can essentially be seen as a change in respective population sizes (with the population being defined by language spoken), with births, deaths, and “migratory” inflows and outflows of people adopting a different language. The balancing equation of population change provides a simple framework for conceptualizing population shifts (Preston, Heuveline, & Guillot 2001).

$$Eq. 1$$

$$N(T) = N(0) + B[0,T] - D[0,T] + I[0,T] - O[0,T]$$

Where

$N(T)$ = number of persons alive in the population at time T,
$N(0)$ = number of persons alive in the population at time 0,
$B[0,T]$ = number of births in the population between time 0 and time T,
$I[0,T]$ = number of in-migrations (here, language switching) between time 0 and time T.
$O[0,T]$ = number of out-migrations (language switching) from the population between time 0 and time T.

Generally speaking, of the demographic processes that affect the vitality of a language group, only the transmission of a language from one generation to another has been considered in the literature. However, mortality and fertility also play vital roles in the future of endangered languages, and are additive with linguistic switching, or changing one's predominant language of communication. In other words, the death of a speaker counts as much against a language's vitality as somebody who ceases speaking the language.

Natural Growth or Decline

Fertility

Endangered languages are found in communities that are at different stages in the demographic transition (or the systematic transition from high death rates and high birth rates to low death rates and low birth rates, Siegel 2018). The importance of transmissibility for maintaining total numbers is inversely proportional to the natural growth rate of the language speakers. For example, in the case of language communities in more developed country contexts such as indigenous American language speakers or the Ainu of Japan, lower fertility rates means that more children need to be taught the language in order for the language to sustain its number of speakers. If the population's mortality and fertility is such that the population is "stationary," which means no growth and no population decline, then the language must be transmitted to every single child in order to avoid linguistic decline. If the language is spoken in a below-replacement fertility population, then the language will decline no matter how effective the language transmission and retention efforts are (assuming no switching into the endangered language). To be more mathematically precise, the transmission rates of the language that are needed for sustainability are related to the Net Reproductive Rate (hereafter NRR). This rate is explained in more detail below, but a NRR of 1 means that a closed population has the fertility and mortality rates to perfectly replace itself each generation. An NRR of 2 means that it is doubling each generation. Mathematically, this means that the rate of retaining each child in the language speaking group needed to maintain stationarity equals $\frac{1}{NRR}$. If a language speaking group has an NRR of 2, then to achieve intergenerational sustainability in terms of raw numbers the language must be transmitted to half (.5) of the children. An NRR of 1.2 means that the community has to achieve intergenerational transmission of at least .83, etc. What this means in practice is that if the endangered language is in a population that exhibits rapid population growth—for example some Sub-Saharan African languages—then the language speaking community can continue to maintain their overall numbers even in the face of children and families switching to other languages. Sub-populations with unique reproductive patterns in developed country contexts can also exhibit this effect even if they are a part of a society that has otherwise lower fertility. Pennsylvania Dutch spoken by the Amish and Yiddish spoken by the Ultra-Orthodox Jewish community (Ryan 2013) are two US examples. However, this is not the case for language speaking communities that are reproducing and surviving closer to replacement level, such as the Ainu or Navajo, in which case they need to retain nearly all of their speakers intergenerationally or else they face the specter of decline.

Mortality

Mortality influences language vitality in several ways. Perhaps most obviously, lower mortality can lead to less rapid decline, especially in cases where older individuals represent a native-speaking past where the language was spoken more exclusively (for example, Ned Maddrell, the last native speaker of Manx, died in his mid-90s in 1974). Therefore, in many cases generic development and public health efforts in developing countries can serve the additional benefit of helping preserve endangered languages.

Mortality rates affect language vitality in other less obvious ways as well. For example, the higher the female mortality rate during the 0-50 age range, the higher the fertility rate has to be to offset losses. On the other hand, even if there is above-replacement fertility and language retention, a larger tail of older individuals from lower mortality rates will lead to lower overall growth rates as they increase the size of the denominator. Growth rates may be misleading, as the generation-to-generation growth factor may paint a different picture.

Population Movements

Finally, migration, here in its conventional demographic sense of people moving through space, may also affect language vitality as it dilutes or concentrates the core of language speakers. However, these effects of population movement on language vitality are distal: they operate through the more proximate mechanisms of the three demographic processes of language switching, births, and deaths.

Data

Many smaller, nationally-representative surveys with Ns in the hundreds or low thousands (for example, the General Social Survey in the United States) may occasionally pick up a handful of speakers of rarer languages and may have more detailed language questions, but surveys of this size will not generally provide enough speakers to provide much useful information about the age structure and characteristics of specific endangered language speaking communities. Obtaining enough information about these communities requires either surveys targeted to the specific groups in question or massive national-level enumerations. In the case of the United States, the annual ACS fully rolled out in 2005 as an intended replacement for the decennial census' long-form survey—is the largest nationwide survey in the United States (with the exception of the decennial census). The ACS has a question on language spoken at home, which serves as a good measure of language use for our purposes. In the ACS, the language data are derived from three questions: "does this person speak a language other than English at home?" and, if yes, "what is this language?" and "how well does this person speak English?"

Even with the sample size in the ACS, the estimates for different demographic parameters are often too imprecise to be meaningful for delineating certain group-specific characteristics such as age structure.

For this study the five year ACS data set for 2013-2017 (an aggregation of the ACS data from 2013-2017) is used. The use of a large survey as opposed to more targeted, group-specific enumerations represents a tradeoff. First, while extremely large national surveys have the advantage of being able to pick up more dispersed speakers that may not be as easily captured, in extremely large-N national surveys, language speaking is operationalized using rather simple single-item measures: there is generally not enough space on the instruments for multi-item measures of proficiency and usage, for example, that are traditionally incorporated into measures of linguistic endangerment (Lee and Way 2016). It is clear that large-N generic surveys do not, by themselves, provide the detail of data required to determine level of endangerment which requires not only dichotomous information on whether they use the language or not, but more contoured variables describing proficiency, context of use, quantity of use, etc.. Basic demographic data should be seen as complementary to more language-specific investigations.

Methods

In this study, I explore several demographic indicators that shed light on language vitality, including net reproductive rate, population momentum, average age, and old-age dependency ratio for the indigenous languages listed earlier. After calculating these indicators for their respective language communities, the different indicators will be conceptually and empirically compared to each other. While many of the principles involved have been theoretically discussed in Siegel (2018), in practice no literature of which I am aware has actually calculated these indicators for smaller language groups, and the practicality of their use in future endangered language scholarship will be discussed.

As noted above, demographic concepts such as aging are often invoked in discussions of endangered language vitality, even though no study has formally used quantitative or demographic methods to measure the concepts invoked. To do so would require some description of the age structure of the respective community. There are a number of measures used in demography to measure and operationalize age structure. However, in each case the pre-verbal (people who do not speak yet) age bracket will have to be imputed from other data if it is to be comparable to general population estimates. In the ACS case, the language questions are only recorded for individuals ages five years and older, so the 0-4 age bracket is derived from the other age brackets. The exact method of imputation is a potential source of variation in estimates. However, given the relatively small proportion of the total population that occupies this age bracket (for example, in 2017 the 0-4 age group constituted 6% of the US population [US Census Bureau 2017]), the potential error introduced is negligible.

In terms of practical outcomes, these demographic measures are intended to provide a more precise sense of the future of their respective languages than would be possible with more qualitative or non-demographic measures, and the potential usefulness of each measure for that end will be analyzed and discussed. Five measures will be discussed here: population pyramids, average age of speakers, dependency ratio, population momentum, and the net reproductive rate. Of these five, all but net reproductive rate will be calculated. Specifically, their practicality in light of data and sample size constraints, their interrelations among each other, and their potential payoff in terms of information gained will be discussed. The analysis and discussion here is not aimed towards producing a “best” measure, but to rather analyze the different benefits, drawbacks, and optimal contexts for each measure. Finally, given the dearth of literature on the indigenous language speaking population, the take-aways for the different groups themselves are also a substantive aim of this paper. The potential usefulness of a cohort-component approach—the standard method of population projection used today—will also be discussed but not calculated, as well as other approaches such as microsimulation. All measures treated here will be compared and contrasted in the discussion section.

Population Pyramid

Simply calculating the population pyramid of each language speaking group would provide the most information. Since some language speaking groups probably have very atypical population structures due to mortality differences, fertility differences, and transmission, any attempt to capture the structure in one number might be misleading. For example, many of the indigenous groups I will discuss below have parabolic age structures, with few young people and few older people, which is atypical. Publishing the raw pyramids would also allow language scholars to incorporate these numbers as they see fit, and population pyramids are quite simple to calculate. Since they are not reducible to a single number, it is more difficult to do time trends (although from a published set of population pyramids the other indicators could be derived by other scholars).

Population pyramids do not by themselves distinguish between decrements from switching and decrements from death; however, reasonable assumptions about trends in mortality can lead to reasonable conjectures about trends in switching. For example, in the case of the Cherokee below there is a significant drop in the percentage of speakers in the 45 and over age cohort compared to that in the 30-45 age cohort (Figure 7; incidentally, this drop chronologically mirrors the drop found in other groups such as the Western Apache and the

Lakota). Absent mortality differentials favoring Cherokee speakers over non-speakers, this drop would be due to generational changes in speaking patterns. Time-trend data would help further distinguish between the lack of adoption of Cherokee by the cohort in childhood and abandonment of Cherokee as people aged.

Average Age of Speakers

This measure is easy to calculate and is the most intuitive. It is also simple to compare across different datasets and censuses, and is easy to track longitudinally to measure changes across time. The downside, as with any single-item measure, is that it necessarily hides much of the variation within the age structure. For example, an average age may be low because of fewer older people due to mortality, or it may be low because of high fertility and reproduction. Both of these scenarios could conceivably lead to the same number, but they present different implications for language vitality.

Dependency Ratios

The old-age dependency ratio of the population is the ratio of the number of 65 and over adults to the number of working-age adults (15-64). This number is often used as a useful way to determine the effects of population aging on a society's ability to support its older population. Like the previous two measures, it provides information about the population structure in a single figure. It is relatively painless to calculate and has the advantage of not requiring imputation of the pre-verbal age group. Consequently, this number, derived for a language group, could be appropriately compared to the same metric for countries or populations as a heuristic for interpretation.

Population Momentum of Speakers

Population momentum has an easily interpretable, substantively important conceptual meaning: by what factor would the population grow before reaching long-term equilibrium if all of its rates (fertility, mortality, and "migration,") immediately shifted to replacement level. For example, a language group, country, or other population with a population momentum of 1 is at replacement level and is neither growing nor declining, whereas a language group with a population momentum of 2, for example, would still increase by double before it settles into an equilibrium if the rate of people starting to learn the language suddenly switched to the exact same rate as people who are ceasing speaking the language (either through language switching or death).

Population momentum is often used to explain the presence of continued population growth even in areas that have below-replacement fertility. In these cases, increasing growth is an artifact of the population structure, and is not due to any above-replacement natural growth or migration rates. Similarly, population momentum has the potential to, in a single figure, describe how much potential growth is baked into its population pyramid. The drawbacks here are methodological: calculating momentums requires a lot of assumptions about what the final population looks like, and they are relatively difficult to calculate. (See Appendix I for technical details on calculating population momentum). Consequently, there are not a lot of published values that can act as standard reference points for comparison. For example, neither the UN nor the World Bank publish tables of population momentums. Finally, population momentum calculations require stable age bracket-specific population estimates which, as will be explored below, are difficult to obtain for smaller language groups. Population momentum calculations are more sensitive to variation within age brackets than other measures that simply collapse age brackets (and therefore have tighter confidence intervals) such as old age dependency ratio.

Net Reproductive Rate

The net reproductive rate (hereafter NRR) is the factor by which the population of interest will grow in a generation given its vital rates. For example, if a population is perfectly stationary, with no growth or decline, it will have an NRR of 1. If it will double in size every generation it will have an NRR of 2. Mathematically, the rate is essentially the total fertility rate modified by the life table risks of mortality to create the number of daughters any given woman could be expected to have if they lived their entire life while being exposed to the age-specific rates of the respective year. If language switching is incorporated into this model by estimating the number of live individuals in each age bracket who start and stop speaking the language, this rate would perhaps most clearly and precisely get at the main item of conceptual interest to endangered language scholars and activists: how much is a language dying or growing with each passing generation? However, this is the measure most difficult to calculate, as it requires not only population size estimates by age bracket, but also age-specific fertility and mortality (and, possibly, switching) rates. Therefore, the population group has to be fairly large for a stable, accurate NRR to be calculated. However, reasonable fertility estimates may be derived by statistically comparing the probability of having a child in the last year for women 15-50 in the respective language group while controlling for age to the baseline population to the same probability for women 15-50 in the baseline population, and then assume that the difference in total fertility rate (TFR) is the same factor difference as the difference in the probability of having a child. For example, if women in group A are twice as likely to have a child in the past year (while controlling for age), then the assumption is warranted that the difference in TFR is also double. (In the case the languages used in this analyses, Inupiaq and Central Yupik show statistically significantly higher fertility than English-only households with odds ratios of a 15-50 year old woman having a child at 2.3 and 1.9, respectively, and Cherokee has significantly lower fertility with an odds ratio of .64 [results not shown]). Similarly, since language-speaking life tables are not generally available, life tables from a comparable population (such as total tribal members in the case of the indigenous American languages) could be used. However, given that these assumed equivalences are all potential sources of error, and that they require extensive external data, NRR estimates will not be derived in this paper, which is instead focused on more readily attainable estimates.

Empirical Comparisons of Different Metrics

With the exception of the NRR, the metrics noted above are calculated and presented for the seven largest indigenous language speaking communities in the 2013-2017 ACS: Inupiaq, Central Yupik, Western Apache, Navajo, Lakota, Eastern Keres, and Cherokee. They are compared in order to examine their consistency and usefulness. Standard errors are calculated using the 80 provided replicate weights at the household and person levels. The 0-5 age bracket is imputed for the momentum figures in order to match the definition of momentum (the factor by which the population would grow if all its rate immediately collapsed to replacement-level), but the age, population pyramids, and old age dependency ratios are calculated only from the speaking population.

Table 1: Average Age, Old Age Dependency Ratio, and Population Momentum of Seven US Indigenous Languages

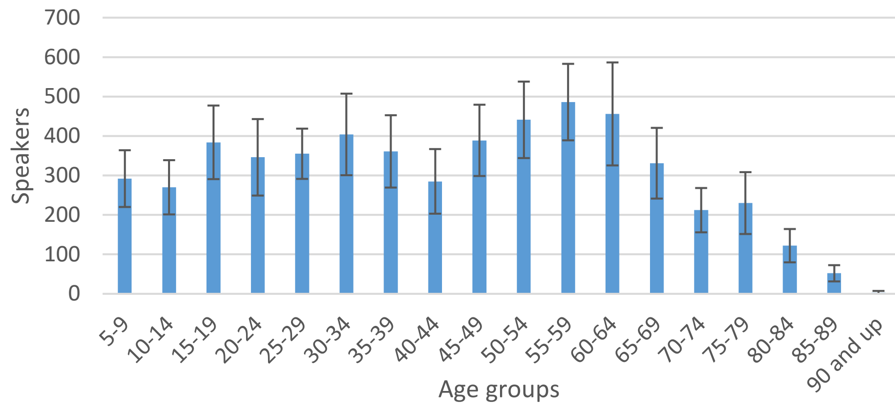
	Average Age (Years)	**Percentage 15+ Working Age (Old Age Dependency Ratio)**	**Momentum**
Inupiaq	43.1 (41.7-44.5)	80% (78%-82%)	.83 (.82-.84)
Central Yupik	38.4 (37.7-39.1)	87% (86%-88%)	1.05 (1.04-1.06)
Western Apache	43.9 (42.9-44.9)	85% (83%-87%)	.68 (.67-.69)
Navajo	42.4 (42.0-42.8)	84% (83%-85%)	.81 (.81-.81)
Lakota	43.1 (41.8-44.4)	83% (81%-86%)	.79 (.78-.80)
Eastern Keres	39.8 (38.7-40.9)	84% (82%-86%)	1.0 (.99-1.01)
Cherokee	44.9 (43.4-46.4)	81% (79%-83%)	.70 (.69-.71)

Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-year data

95 percent confidence intervals are placed after the calculated figures.

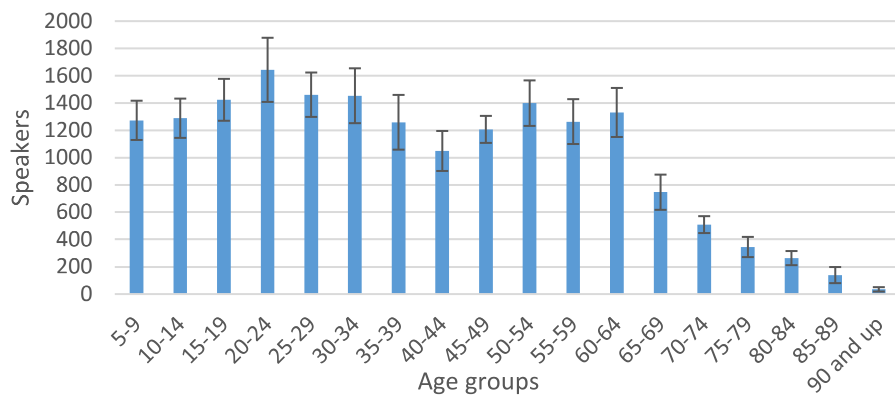
For more information on the ACS, see census.gov/acs.

Figure 1: Population Pyramid of US Inupiaq Speakers by Age



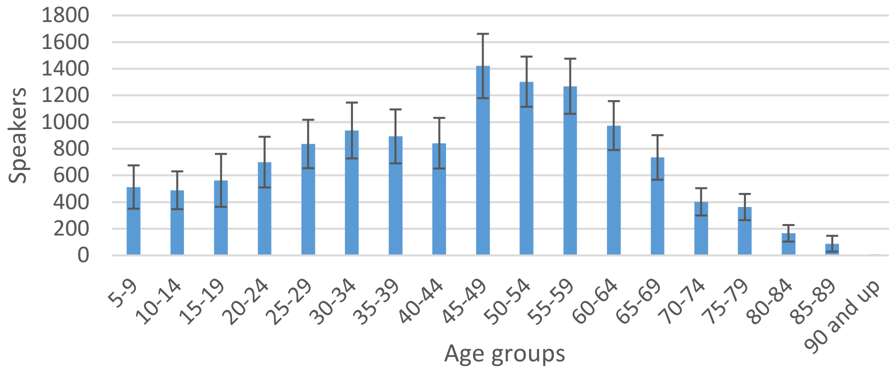
Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-year data.
Bars represent estimates and 95% confidence intervals.
For more information on the ACS, see census.gov/acs.

Figure 2: Population Pyramid of US Central Yupik Speakers by Age



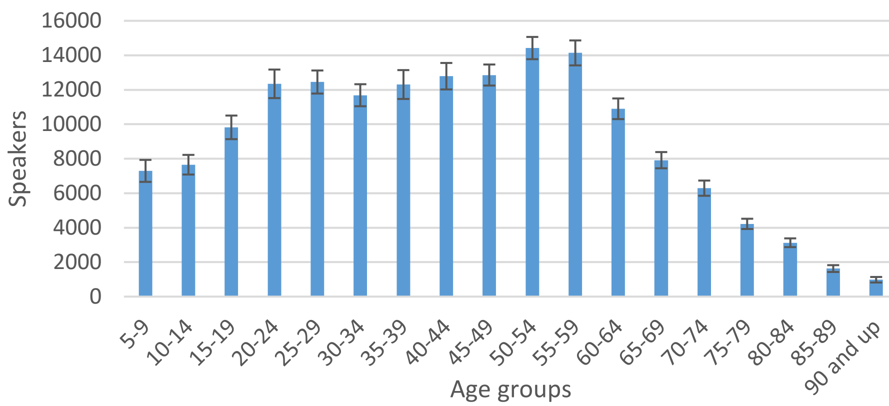
Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-year data.
Bars represent estimates and 95% confidence intervals.
For more information on the ACS, see census.gov/acs.

Figure 3: Population Pyramid of US Western Apache Speakers by Age



Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-year data.
Bars represent estimates and 95% confidence intervals.
For more information on the ACS, see census.gov/acs.

Figure 4: Population Pyramid of US Navajo Speakers by Age



Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-year data.
Bars represent estimates and 95% confidence intervals.
For more information on the ACS, see census.gov/acs.

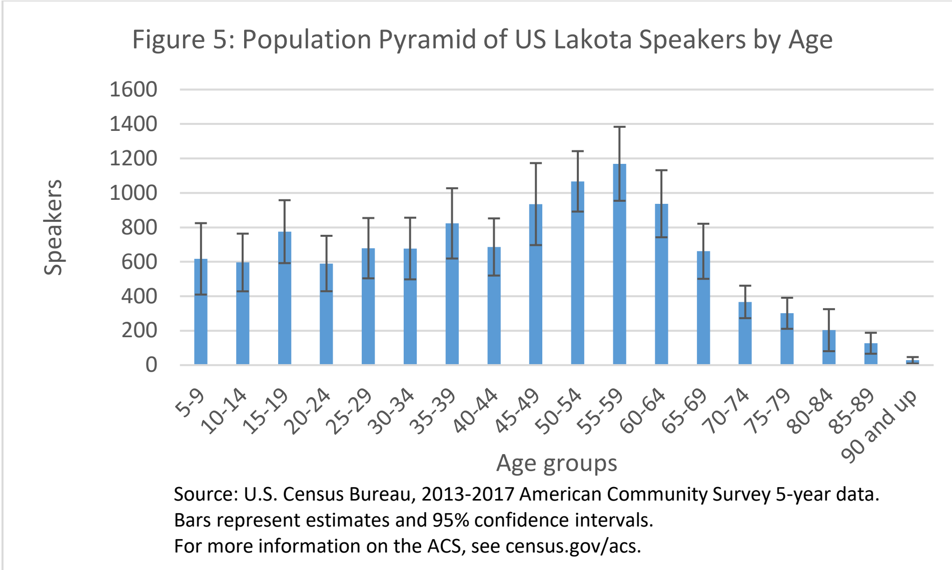


Figure 5: Population Pyramid of US Lakota Speakers by Age
1600
1400
1200
1000
800
600
400
200
0
Speakers
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85-89
90 and up
Age groups
Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-year data.
Bars represent estimates and 95% confidence intervals.
For more information on the ACS, see census.gov/acs.

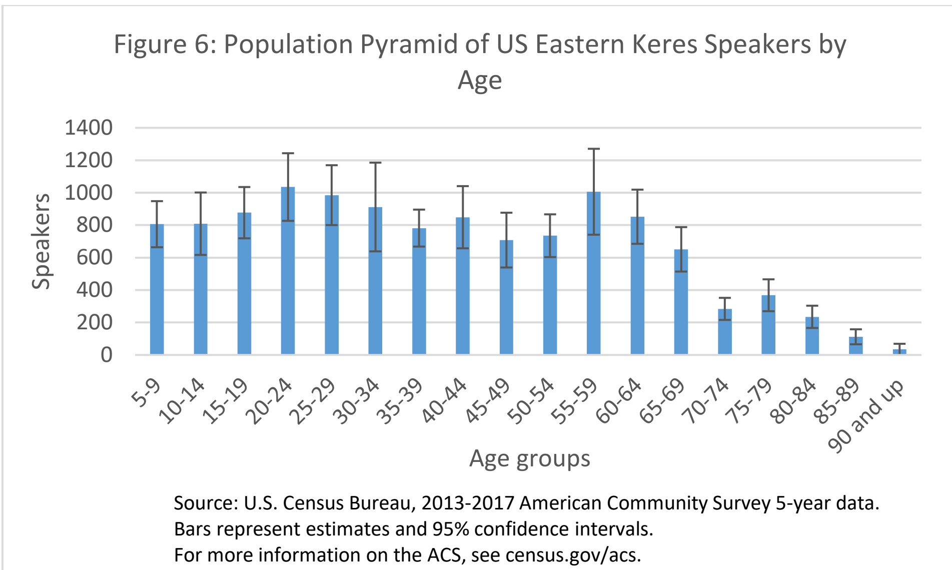
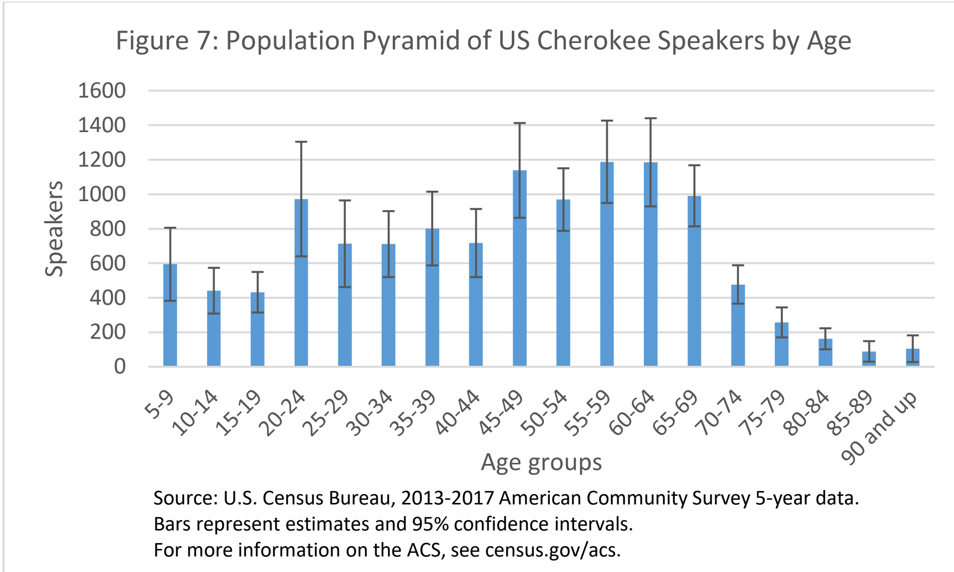


Figure 6: Population Pyramid of US Eastern Keres Speakers by Age
1400
1200
1000
800
600
400
200
0
Speakers
5-9
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85-89
90 and up
Age groups
Source: U.S. Census Bureau, 2013-2017 American Community Survey 5-year data.
Bars represent estimates and 95% confidence intervals.
For more information on the ACS, see census.gov/acs.



DISCUSSION

As noted above, different measures of population structure can paint different pictures. For example, many of the old-age dependency ratios lie in a similar mid-80s range, even when the shape of the population pyramids vary. However, for the most part these measures all point in the same direction, with the more top-heavy population pyramids (indicating older populations) yielding smaller population momentums and higher average ages. This means that the general picture in terms of long-term vitality is not sensitive to the choice of indicator used.

Nearly all the languages examined here show evidence of decline across cohorts (Figures 1, 3, 4, 5, and 7). The two possible exceptions are Central Yupik (Figure 2), which shows a population momentum score above 1 at 1.05, and Eastern Keres, which shows approximately replacement-level rates at 1 (Figure 6).

Even with five-year ACS data—the largest current survey of these languages possible in the US context—some of the groups have relatively large confidence intervals for their age brackets, suggesting that the demographic estimates themselves are not very reliable. Consequently, traditional demographic measures of population structure are most useful for relatively large minority language groups such as the Navajo or Central Yupik. For these larger groups, the "gold standard" of population projection is a cohort-component approach. In addition to their well-known use for projecting future country-level populations cohort component projections have been used to project other demographics such as religious and political affiliations (Kaufmann, Goujon, & Skirbekk 2012; Skirbekk, Kaufmann, and Goujon 2010). However, conducting a proper cohort component projection requires estimates for age-specific mortality, age-specific fertility, and age-specific switching rates (traditionally using five-year age intervals), so a proper projection would require not only a tabulation of births and deaths, but also a year-to-year tabulation of how many people in each age bracket switched languages. Such data are unavailable using the ACS or other large-scale surveys.

Even in cases where a complete census of all language speakers for a group is taken, in the case of smaller groups stochastic population processes may add too much noise for the estimates to be very useful. One possible method of modelling language futures for extremely small, critically endangered groups (e.g. those with speakers in the dozens or fewer) is agent-based microsimulation. There are several user-friendly open source agent-based modeling programs such as Netlogo that can be used for such a purpose. When this approach is taken, the individual "agents" involved can be assigned varying levels of parameters such as linguistic ability, chance of passing on the language to children, fertility rate, mortality rate, etc. and the simulation can be run multiple times under different parameters. This way, exactly what kind of forces have to be at play for the language to reach sustainability or growth can be estimated.

CONCLUSION

Demographic methods and approaches have the potential to clarify and add precision to the study of language vitality. However, the more sophisticated and precise the approach, the larger the sample size and more comprehensive the information is needed in order to calculate estimates. In the absence of information on age-specific linguistic switching, mortality, or fertility rates, the information from the age structure is helpful in fleshing out the general future picture of the respective language. Furthermore, these estimates appear to be strongly related, suggesting that simple measures such as average age may provide a simple heuristic for determining demographic vulnerability. Additionally, simple population pyramids (with confidence intervals) may be effective at simply visually sketching out the general future trends for language population, with bottom-heavy pyramids yielding indicators that denote growth, with top-heavy pyramids yielding indicators that denote contraction. For extremely small endangered language populations, microsimulations using a variety of inputs for mortality, fertility, and transmission may provide another approach to predicting their future trajectory.

APPENDIX

Calculation of population momentum requires three derived distributions (Preston, Heuveline, & Guillot 2001, 161-7). The first is *cs(a)*, the proportionate age distribution of the eventual stationary population that will form once the replacement-level vital rates have had enough time to work their way through the age structure of the population. While mortality rates undoubtedly vary from indigenous group to indigenous group, the final stationary population is itself a hypothetical construct that forms after rates have changed; therefore, the most recent publicly available US life table (from the 2014 National Vital Statistics Report) is used to construct *cs(a)*. This choice also has implications for the conceptual interpretation of the population momentum figure: the population momentum represents the proportion increase or decrease of the population if mortality rates were lowered to match those of the baseline US population *and* fertility rates were commensurately changed to replacement level. Therefore, the numbers derived from these rates are meant to construct a standardized hypothetical measure rather than as a prediction for the future. This standard measure also lends itself to an easily comprehensible conceptual interpretation: how much the language-speaking population would grow if its mortality and fertility rates immediately changed to replacement-level (which would either be an increase or a decrease).

Second, the weight component—*w(a)*—is derived from taking the expected lifetime births above age *a* as the numerator with the mean age at birth in the stationary population as the denominator. In practice the shape of this function does not vary significantly from location to location even in cases of significant differences in fertility and mortality regimens. (Again, here "births" are assumed to occur at biological birth, although a birth into a language group can also be conceptualized as learning the language). Here once again the latest age-specific fertility rates for the US are used (2015), and are comparable to other published rates. The *w(a)* and *cs(a)* distributions represent baseline components that do not vary from case to case; the *c(a)* distribution is the actual proportionate age distribution of the language population.

After these three terms are derived they are used to calculate the estimated population momentum using equation 1.

$$Eq. 2$$

$$\sum w(a) * c(a)/cs(a)$$

REFERENCES

Abrams, Daniel. M., and Steven H. Strogatz, (2003). Linguistics: modelling the dynamics of language death. *Nature* 424(6951): 900.

Kaufmann, Eric, Anne Goujon, and Vegard Skirbekk. (2012). American political affiliation, 2003–43: A cohort component projection. *Population studies* 66(1): 53-67.

Kenneth L. Regh and Lyle Campbell (editors) (2018). *The Oxford handbook of endangered languages.* Oxford: Oxford University Press.

Nala Huiying Lee and John Van Way. (2016). Assessing levels of endangerment in the catalogue of endangered langauges (ElCat) using the language endangerment index (LEI). *Language in Society* 45: 271-292.

Preston, Samuel, Heuveline, Patrick., and Guillot, Michel. (2001). *Demography: Measuring and modeling population processes*. Oxford, UK: Blackwell Publishers.

Ryan, Camille (2013). Language use in the United States. (2011). Census Bureau Report ACS-22. Suitland, MD. https://www.census.gov/library/publications/2013/acs/acs-22.html.

Siegel, Jacob S. (2018). *Demographic and socioeconomic basis of ethnolinguistics*. Cham, Switzerland: Springer.

Skirbekk, Vegard, Eric Kaufmann, and Anne Goujon. (2010)"Secularism, fundamentalism, or Catholicism? The religious composition of the United States to 2043." *Journal for the Scientific Study of Religion* 49(2): 293-310.

UNESCO. Language Vitality and Endangerment. (2003). UNESCO Ad Hoc expert group on endangered languages. International expert meeting on UNESCO. Paris, 10–12 March 2003.

US Census Bureau. Generated by BLINDED FOR REVIEW using American FactFinder; <http://factfinder.census.gov>; (1 December 2018).