Introductionludwig the language that an individual speaks truly

INTRODUCTIONLudwig Wittgenstein once said that " The limits of my language mean the limits of my world." The diverse spectrum of languages that circulate the globe is what makes our world interesting, multi-faceted, and beautiful. The language that an individual speaks truly impacts and defines the unique way he perceives and understands the world around him —no matter the subject or discipline. I have chosen to research the trends from United States' Census data related to non-English household language use because I feel that the circulation of languages will continue to make the world, as Wittgenstein phrases it, less " limited," and more open to progressive ideas expressed through multi-lingual dialogue between people of different backgrounds.

Specifically, I will investigate the trends of Chinese, French, Italian, Russian, and Korean spoken in the home. During my time at Stanford this past summer, I found it interestingly surprising that many of my classmates were either from China or spoke Chinese fluently in their homes. As the globalization of China continues, I predict that its widespread influence will demonstrate to a noticeable increase in the number of people who speak Chinese in the home in the upcoming 2020 U. S. Census. In addition, I chose French and Russian because I take IB French and I am also independently studying Russian in my spare time.

Because I would like to pursue a career in global studies, both languages are directly pointed toward my European diplomatic interests, and I hope for the sake of preservation that both languages are increasing in U. S. households in the near—and distant—future. Next, I decided to investigate Spanish spoken in U. S. households because my mother used to be an ESL (English as

a Second Language) teacher for many years, and through experience with her students, I saw how many families from Spanish-speaking nations were immigrating to our country.

Finally, I chose to look at the Korean language because my best friend recently spent the summer in Seoul, and exposed me to the relevancy of the Koreas on an international scale. While the other four languages appear in lists on countless websites that predict the top ten most influential or powerful languages in the future, Korean does not appear on these lists. I think it will be interesting to compare the trends I predict of a language that is not expected to have a large global presence with others that will.   In this exploration, I will gather data from several decades of U.

S. Census records detailing how many people over the age of five (who were living in the U. S.

) spoke a language other than English between the decades of 1980 to 2010 to predict the trends of these languages spoken in U. S. households in future Censuses.

While the United States is only a small portion of the world's population, I think that it can be a prime identifier for any global fluctuations in these five languages over time because the " melting pot" nature of our country is a good representation of the percentages of languages spoken internationally. I will be exploring the trends of the populations of the United States over the ages of five who speak Chinese, French, Spanish, Korean, and Russian at home with the r2 correlation coefficient test and bar graphs to calculate the predicting trends of the total speakers residing in the U. S. and over the age

of five of each language by 2030, 2050, and even 2100. Through such statistical analyses, I hope to learn how closely the trends of the five languages resemble one another represented in the national census data, or if further investigation may suggest that they will vary greatly in decades to come.   BACKGROUND   Before beginning this investigation, it is useful to know some information regarding the immigration of the peoples who are responsible for the presence of these five languages in the United States. The graph below from the Migration Policy Institute shows the influx of the total immigrant populations into the United States from 1850 to 2009. It is important to keep in mind for this investigation that when foreign-born immigrants assimilate into a new country, they bring their native languages with them, altering that country's overall language demographic.

Making up approximately the third-largest immigrant population in the country, former Chinese citizens arrived in the U. S. looking for employment in two waves: the first during the mid-19th century, and the second lasting from the 1970s to the present day. Economic difficulties in China resulting from its defeat during the Opium Wars against England during the early 1840s was the main cause of the first migration wave, while the second was due to the 1965 Immigration Act (which allowed non-European immigrants to enter the United States again), " China's loosening of its emigration controls in 1978, and the normalization of U. S.

-China relations in 1979″ (U. S. Immigration Trends). The largest concentrated populations of Chinese immigrants reside in the states of New York City and California, where the influx of Chinese immigrants along both coasts of the country a few hundred years ago led to the creation of large

communities that continue to maintain Chinese culture and language today. Graphic from: https://www.

migrationpolicy. org/article/chinese-immigrants-united-statesThe first major influx of French immigrants began in 1846, attempting to escape the death and poverty from the potato famine that was not only affecting Ireland, but also most of Europe (Alchin). Many more French immigrants continued to pour in throughout the 1840s and 50s due from fear of Napoleon III's dictatorship (Alchin). However, as the French government began to stabilize, fewer French citizens immigrated to the United States. Today, large French-speaking populations reside in cities like New Orléans, Lousiana, where Cajun settlers from Canada established communities.

There are many French speakers in Maine due to its proximity to the Canadian border.   The first significant mass-arrival of Spanish-speaking immigrants from Latin America occurred simultaneously with that of the French immigrants in the mid-1840s, dramatically increasing again during the first three decades of the 20th century. The California Gold Rush and constant change-over of land between the U.

S. and Mexico were the motives for migration. From islands like Puerto Rico and Cuba, another influx of immigrants occurred during the 1960s and 70s due to political unrest and danger. Since then, the Spanish-speaking population of immigrants continues to increase exponentially due to gang violence, governmental uncertainties, and poor economies. Today, " about 58 percent of the estimated total in 2010" of immigrants come from Mexico alone (Gutierrez).

According to a Pew Research Center survey from 2011, The most Spanish-speaking immigrants live in the states of Texas, Florida, and California (Brown). Sanctions and geographical difficulties kept many Russian immigrants from entering the U. S. throughout much of the 19th century. However, in the 1880s and 90s, immense famines and cholera caused many Russians to flee to the U.

S. The number of Russian immigrants continued to increase into the 20th century and through the beginning of the first world war. The chaos of the Russian Revolution and the 1911 Dillingham Commission Report, stating that the U. S. would put priority on " Old Immigrants" hailing from Western Europe, rather than " New Immigrants" from Eastern Europe and Asia, caused migration numbers to drastically decline (Alchin " Russian Immigration to America"). Continual death, political opposition, and economic difficulties resulted in few Russians moving to the United States afterward. Today, most Russian-speaking populations are concentrated in the states of New York and California. Few Korean immigrants came to the U.

S. until after the Korean War and the removal of the United States' " Immigration Removal Act of 1965" (U. S. Immigration). Starting in the 1960s, the Korean immigrant population exponentially increased until the 2010s.

Today, approximately half of the Korean immigrant population resides in the states of California, New York, and New Jersey (U. S. Immigration). The (MPI) demonstrates the recent trend of the " Korean immigrant population in the United States becoming stagnant.

As economic and political conditions in South Korea have improved, fewer people have been interested in emigrating" (U. S. Immigration). Now, whether the total number of Korean immigrants coming into the U. S. accounts for immigrants from both South and North Korea, or South Korea alone, I am not sure. However, it seems most likely that the statistics encapsulate South Korean immigrants because North Koreans have little to no freedom or opportunities to escape their country.   Graphic from: https://www.

migrationpolicy. org/article/korean-immigrants-united-statesIn order to do this investigation, you need a solid understanding of what the values represented in the U. S. Census data demonstrate. To grasp the Census' statistics, I examined corresponding data from the Migration Policy Institute between the years of 2009 to 2015:   http://www. migrationpolicy. org/programs/data-hub/us-immigration-trends#lepAccording to the Migration Policy Institute's data from 2009 to 2013, I learned that the most common language spoken by citizens (ages 5 and older) was Spanish, with 37, 459, 000 speakers, 44% of whom were Limited English Proficient (LEP), meaning that their comprehension or speech in English was severely limited.

Chinese ranked 2nd, with 2, 897, 000 speakers, 55% of whom were (LEP). French ranked 5th, with 1, 308, 000 speakers, 20% of whom were (LEP). Korean ranked 6th, with 1, 117, 000 speakers, 55% of whom were (LEP). Russian ranked 11th, with 879, 000 speakers, 47% of whom were (LEP). Graphic from: http://www. migrationpolicy. org/programs/data-hub/us-immigration-trends#lepBy 2015, Spanish, Chinese, and French maintained the same rankings, but the number of Korean speakers decreased and the

number of Russian speakers increased. Between the data spanning the years of 2009-2013 and the data from 2015: The total number of Spanish speakers increased by approximately 2, 587, 000, and the percentage of (LEP) speakers decreased by 3%The total number of Chinese speakers increased by approximately 437, 000, and the percentage of (LEP) speakers increased by 0.

7%The total number of French speakers decreased by approximately 42, 000, and the percentage of (LEP) speakers increased by 0. 1%The total number of Korean speakers decreased by approximately 8, 000, and the percentage of (LEP) speakers decreased by 1. 8%The total number of Russian speakers increased by approximately 26, 000, and the percentage of (LEP) speakers decreased by 3%These trends I calculated before my actual analysis show that the number of Spanish, Chinese, and Russian speakers (whether or not they spoke fluent English) increased, while the number of French and Korean speakers decreased. It is also interesting to note that the percentage of Spanish, Russian, and Korean speakers who were (LEP) speakers decreased considerably, while the percentage of Chinese and French speakers who were (LEP) speakers increased by very small margins. This data shows us two things: the first being that it seems that more Spanish, Russian, and Chinese native speakers are moving to the United States every year, while the number of French and Korean speakers is decreasing, and secondly, that even though it seems that less Korean speakers are entering the country each year, the percentage of Korean (LEP) speakers is lessening.

Another trend suggests that a larger percentage of the Russian- and Spanish-speaking populations are proficient English-speakers upon survey. Contrastingly, French-speakers and Chinese-speakers seem to have decreasing percentages of their populations who speak English proficiently upon being surveyed.   EQUATIONS AND PROCESSES TO KNOWTo successfully complete this mathematical investigation, one must be familiar with correlation, Pearson's correlation coefficient, r (Haese 550) and the coefficient of determination, r2 (" Correlation Coefficient"). 1. Correlation is defined as the " relationship or association between two variables" (math bk pg 546). Looking at plotted data, one might notice downward (or negative) trends, upward (or positive) trends, or no correlation in the data set at all (Haese 547). In addition, one might also look at how linear a trend is, and whether its points model a strong, moderate, or weak correlation based on the number of points that fall on—or extremely close to—the trendline (Haese 547).

Correlation, as well as the terms associated with describing relationships between variables in data sets, will be extremely important to help understand, define, categorize, compare, and contrast the different data sets of the investigation. 2. Pearson's correlation coefficient, r, is an equation to determine a " more precise measure of the strength of linear correlation between two variables" because using correlation alone can be inaccurate (Haese 550). As it is important to understand the math by hand before using a calculator to compute the r value, this is the formula of the Pearson's correlation coefficient that I will be using: r=?(x-x')(y-y')?(x-x')2 ?(y-y')2KEY: = sum of all data valuesx and y = the ordered pair points given/calculatedx'

= the mean of the x datay' = the mean of the y dataIt is also important to understand that the r value ranges from +1 to -1, and the " sign" of r shows what the " direction" of the correlation will be, while the " size" of the r value demonstrates how " strong" the correlation will be (Haese 551). A " perfect" negative correlation would have an r value of -1, while a " perfect" positive correlation would have an r value of +1 (Haese 551). 3. The coefficient of determination, r^2, is extremely helpful for this investigation as well, because it is a means " that allows us to determine how certain one can be in making predictions from a certain model/graph" (" Correlation Coefficient"). While the r value shows the direction and strength of a correlation on a graph, the r^2 value, ranging between 0 to +1, " represents the percent of the data that is the closest to the line of best fit" (" Correlation").

Furthermore, I will use the r^2 value taken from multiplying the Pearson's correlation coefficient, r, by itself to determine which types of graphs (whether they be linear, quadratic, exponential, etc.) best fit the data trends that I predict and those that I analyze from past Census data collections (" Correlation"). THE INVESTIGATIONThe first data set that I evaluated came directly from the U. S. Census Bureau chart depicting the " Languages Spoken at Home for the Population 5 Years and Over: 1980, 1990, 2000, and 2010." From this chart, I specifically looked at the populations (" 5 Years and Over") who were speaking Spanish, French, Russian, Chinese, and Korean at home instead of English.

To start, I calculated the linear model's r value, and then the exponential, quadratic, cubic, logarithmic, and quartic models' r and r^2 values of the data set regarding the speakers of each individual language separately. I

determined the r value of the linear model of the data and the r2 values of exponential, quadratic, cubic, logarithmic, and quartic models of the data to see which type of graph was most ideal for the language trends I saw emerging. Below, I have illustrated how I determined the r value of data regarding the number of Russian speakers over the age of 5 who speak Russian in U. S. homes as their primary language. For this example, I show how to determine the r value of the linear model of this data. To do this, I first set up a table to calculate the sums and means of my x and y coordinates: Year: xy(x-x')(y-y')(x-x')?(y-y')(x-x')2(y-y')2(1980)3173226-1. 9-397905756019.

53. 611. 58328389E11(1990)4241798-0. 9-329333296399. 70. 811. 084602249E11(2000)57062420.

113511113511. 10. 011. 825498232E10(2010)68549551. 1283824312206.

41. 218. 055606298E10(2015)6.

58794341. 6308303493284. 82. 569. 505073981E10TOTALS: 24. 528556551871421.

58. 24. 60650399E11AVERAGES: 4. 9571131 I choose to label my x values (years) as 3, 4, 5, 6, and 6. 5 because later on in the investigation, I will also be using data beginning in 1960, which will be year " 1." My y values are from the Census data counts. After finding the averages and totals of my x- and y-values, I took each x-value and subtracted the average x-value from it.

I did this for my y-values as well. Next, I multiplied the (x-x') column values with the (y-y') column values. Then, I squared the (x-x') column values and I

squared the (y-y') column values. After obtaining these values, I then plugged in the values into the r equation: r =?(x-x')(y-y')?(x-x')2 ?(y-y')2 = ?(1871421. 5)?(8. 2) ?(4.

60650399E11)Once I solve the equation above, I get the r value to be 0. 9628950687. This number is very close to the ideal +1, so this demonstrates that this range of data would be fairly accurately represented with a linear model.

Using the data points from the Census data and my calculated r value and r2 values of linear, quadratic, exponential, cubic, logarithmic, and quartic models of each language, I graphed the five languages separately with the purpose of looking for interesting trends among them: I chose to display the trendline of each graph based on which type of graph seemed to fit closest to the ideal "+1" r2 value. When I only looked at a trendline based on the five data points of each of the five language graphs, the Quartic type of function demonstrated an r2 of +1 for all five graphs. While I initially assumed that my five graphs would each uniquely have a different model function type that would be the most accurate trendline, my resulting five quartic best-fitting graphs (via my r2 calculations) debunked my hypothesis.

However, because the amount of data is fairly limited to only five points, I plotted all six types of graphs together and zoomed my graphs out to assess whether a Quartic function would truly be the best-fitting trendline for my data over a longer, more predictive period of time. Interestingly enough, even though the Cubic and Quartic types of functions demonstrated the most ideal r2 values of respectively 0. 999757 and 1 for the graph modeling

the number of primary Spanish speakers, once I zoomed out of the frame, I noticed that the both types of functions immediately decreased at a rapid rate, suggesting that if a Cubic or Quartic model of data fit the trend of Spanish speakers, that Spanish would " die out" completely from U.

S. homes as a primary language by the year 2040! Based on the history and data following Spanish's influence in the world, the possibility of Spanish completely disappearing as a primary language from U. S. households in approximately 20 years seems extremely improbable. Meanwhile, once I looked at the table values of all six graphs compared side-by-side, the other four types of graphs all suggest that there will continue to be a rise in the number of Spanish speakers in U. S. homes by 2050.

The Quartic model of graph is best-fitting for the span of years from 1980 to 2015, but it is not best-fitting when it comes to predicting the future of the language. Because the Linear function modeling the Spanish data had the third-most ideal r2 value of 0. 99076, I believe that it is the best way to predict the trends of how many people in the United States (ages 5 and over) speak Spanish as their first language, rather than English. According to the Linear trendline, there will be approximately 51422086. 8 primary Spanish speakers in the U. S. by 2030; by 2050, 67692267. 2; and by 2100, 108367718.

2. LANGUAGE: BEST-FITTING FUNCTION TYPE: YEAR 2030: YEAR 2050: YEAR 2100: SpanishLinear51422086. 867692267.

2108367718. 2FrenchRussianChineseKoreanAT THE END TALK ABOUT HOW LINEAR PROBABLY WONT HAPPEN BC CARRYING CAPACITY OF EARTH + POP

INCREASE After determining which types of functions fit the regression (trend) lines of the five languages, I decided to compare my findings to that of the Migration Policy Institute's data on total immigrant populations in the United States from specific countries who migrated to the United States between 1960 and 2015. The purpose of using this data was twofold: to 1) have access to a larger range of data and 2) to roughly look at the similarities or differences between the graphs that show the influx of immigrants from nations where the five languages are prominent and the graphs that show the number of speakers over the age of 5 who speak these five languages as their primary languages at home.   For the recorded immigrant population graphs, once again, the Quartic model appeared to be the " best-fitting" trendline for four of the five graphs.

The only immigrant population that was not best-fit by a Quartic model was the graph depicting Russian immigrants because there is no data available from 1960 to 1990, so a Quadratic model fits this data " best." This can probably be attributed to the animosity between the United States and Russia during the Cold War period beginning in the late 1960s, and not fully ending until the early 2000s. Either the United States had blockades in place to keep Russian immigrants out during this time, Russia would not let them leave, or any immigrants who came here from Russia kept quiet and evaded federal surveys like the U.

S. Census. This data is similar to the previous graph set depicting the five spoken languages because both (for the most part) suggest that a Quartic model is " best-fitting," but once the graphs are " zoomed" out to show a

larger period of time confirming that a Quartic model is not the best-fitting trendline available.

https://www. census. gov/prod/2013pubs/acs-22. pdf Work CitedAlchin, Linda. " French Immigration to America." French Immigration to America: a History for kids ***, www. emmigration.

info/french-immigration-to-america. htm. Accessed 12 Oct. 2017.

Alchin, Linda. " Russian Immigration to America." Russian Immigration to America: History for kids ***, www.

emmigration. info/russian-immigration-to-america. htm. Accessed 12 Oct. 2017. Brown, Anna, and Mark H.

Lopez. " II. Ranking Latino Populations in the States." II. Ranking Latino Populations in the States | Pew Research Center, Pew Research Center, 29 Aug. 2013, www.

pewhispanic. org/2013/08/29/ii-ranking-latino-populations-in-the-states/. Accessed 7 Oct. 2017. " Correlation Coefficient.

" Statistics 2 – Correlation Coefficient and Coefficient of Determination, MathBits, https://mathbits. com/MathBits/TISection/Statistics2/correlation. htm. Accessed 21 Oct. 2017. Gutierrez, David G.

" American Latino Theme Study ." An Historic Overview of Latino Immigration and the Demographic Transformation of the United States,

National Park Service U. S. Department of the Interior, https://www. nps. gov/heritageinitiatives/latino/latinothemestudy/immigration. htm#_edn1.

Accessed 8 Oct. 2017. Haese, Robert, Sandra Haese, Michael Haese, Marjut Maenpaa, and Mark Humphries. Mathematics for the international student Mathematics SL for use with IB Diploma Programme. 3rd ed., Marleston, Haese Mathematics, 2013, pp. 546-55.

U. S. Immigration Trends | migrationpolicy. org, Migration Policy Institute, 2017, https://www. migrationpolicy.

org/programs/data-hub/us-immigration-trends#lep. Accessed 3 Oct. 2017.

R^2: STAT, CALC, #4, 5, 6, 7 (whichever graph you want to try), Calculate, the closer the r^2 value is to 1, the better the values fit with the graph typeTo edit your lists, STAT, EDITtalk about in conclusion about why initial ideas of sequences don't workstate that you need more datatalk about how education is including more English in other countries now—relate that to the results you see todaytalk about what might happen in the future discuss the specifics about the Spanish speaking population—where are they from?