Contents

1 Introduction

The world has changed at an incredible rate over the last few decades. The post-modern world brings along with it rapid shifts on a grand scale over a variety of aspects, including linguistic demography. Language is becoming increasingly important to survive in a "flat" world where each person is only one click away from the remaining 7 billion. The processes of globalization and modernization renders language a critical requirement for worldwide communication. Companies and organizations strive to become more global and well integrated with the rest of the globe. The need arises for a method to rate the power and the usability of the 7000+ languages around the world for purposes of communication and future planning [1].
Various factors affect the changes in language uses. Shifts in the world's demography including population growth and international migration have a direct causal relationship with the distribution of language users. The acquisition of a language allows for access to the wide array of opportunities from countries and communities to which the language is native. By analyzing the trends of global economic dynamics, we can determine the utility of a language and predict its potential in the longer run. In this paper, we propose the use of a system of differential equations to model the rate of change of the distribution of language usage around the world over a period time. This model will take into account the world's population growth, the utility of languages based on economic factors and projections of future migration trends.

2 Language Distribution Model

2.1 Assumptions

Our model is designed based on a set of basic assumptions that allow us to simplify the model and focus our attention on several important factors such as population growth, migration trends, and the world's economy. We will model the distribution of languages around the world based on these main factors.

1. The net changes in the number of a language's users are directly proportional to the net changes in the populations of countries where the language is spoken. This assumption is based on another assumption that the entire population of a country will be able to use its official language(s) either as a

native speaker (L1) or a second-language speaker (L2).[1]

2. To estimate the trends of language distribution in the future, we propose an index to measure the language's share in the global market, called the language utility. This language utility is calculated based on the economic statistics of all the nations which recognize this language as an official language (such as gross domestic product (GDP), GDP per capita, GDP growth rate. This utility will be used to estimate the global migration trend in the future.

3. The total number of annual international migrants (migrant stock) is proportional to the world population. Using this assumption, we project the future migration rates. Other factors that affect the migration stock (e.g. climate change, political stability, environment) are considered as constants because they require more complicated research and data to model.

4. Any immigrant will be L2 speaker of the native language(s) of their destination.

5. Immigration trends will be predominantly from developing countries to more developed ones, and from countries with lower economic growth rates to faster ones.[2]

6. We only consider the current top 14 languages with the most number of speakers in this model: Mandarin Chinese, English, Hindustani (Hindi, Urdu), Spanish, Arabic, Malay (including Indonesian and Malaysian, Russian, Bengali, Portuguese, French, Hausa, Punjabi, Japanese, German).[1] These languages are spoken in economies that have the most potential to grow, or whose population growth rates are expanding considerably. Any

[1]This assumption excludes the cases of India, China, Canada, the Democratic Republic of Congo, Nigeria, Madagascar, Ivory Coast, and Cameroon, whose language distributions are too complicated and populations too significant. India has two official languages: Hindi and English [2]. It is the second most populous country in the world, with an estimated population of 1,324,171,354 in 2016 [3]. Chinese has several different dialects, but the major and official one (the one we consider) is Mandarin. Based on this assumption, all of China's population will speak Mandarin, but this is not true in reality. In some local areas, people do not necessarily use Mandarin. This applies for the rest of the countries in the list of exceptions.

[2]In its 2017 Migration Report, the United Nations evaluates that "most of the growth in the global population of international migrants has been caused by movements toward high-income countries"[4].

shifts to these languages will most likely alter the current distribution of languages around the world.

2.2 Factors of Influence

We identify the following factors as the most important variables in our model and also the most influential ones to predict future linguistic trends:

1. The world population (see section 2.1.3)
2. Countries' populations (see section 2.1.1)
3. Countries' *de jure* languages (see section 2.1.1)
4. Migration trends (see section 2.1.4)
5. Global economic indices (see section 2.1.3)
6. Other factors such as the environment and the local political stability will be treated as constants (see section 2.1.3)

2.3 Model for Language Distribution

2.3.1 Variables

P: population (million people)
t: period of time (year)
$\frac{dP}{dt}$: population growth rate (million people/year)
k: rate of growth
K: capability of the environment (million people)
M: global migrant stock (total number of migrants) (million people)
$\frac{dM}{dt}$: the rate of change of the global migrant stock over a period of time (million people/year)
P_m: language m's population (including both L1 and L2 speakers) (million people)
k_m: language m's population growth rate
K_m: capability of the environment of all the countries that speak language m (million people)
U_m: language utility (the usability of a language. This index represents the share of this language in the global market's language usage and is calculated based on GDP, GDP growth rate, and GDP per capita) (%) ($0 < U_m < 1$)

$L_{1,m}$: the number of language m's native speakers (million people)
$\frac{dL_{1,m}}{dt}$: the rate of change of the number of native speakers of language m over time (million people/year)

2.3.2 Logistic Growth Model

Let $\frac{dP}{dt}$ be the growth rate of the population. Using the logistic growth model, existing data of world population and numerical solving for differential equation, we can calculate the constants k and K using this function[12]:

$$\frac{dP}{dt} = kP\left(1 - \frac{P}{K}\right)$$

2.3.3 Main Model

Using a similar approach, we can calculate the growth rate of a language's population. The usage of a language is dependent on the populations of all the countries that use it as an official language (see section 2.2.1). In addition, we also take into account the rate of change of the migrant stock. Considering the language utility, we estimate the rate of change of the number of immigrants to these countries. The higher the utility, the more attractive the language is economically. Since it reflects the share of the language in global usage, the product of U_m and $\frac{dM}{dt}$ gives an estimate of future migration trends.

$$\frac{dP_m}{dt} = k_m Pm\left(1 - \frac{P_m}{K_m}\right) + U_m\frac{dM}{dt}$$

Applying the logistic growth model to native speakers, we can find the growth of L1 speakers over a period of time:

$$\frac{dL_1, m}{dt} = k_m L_{1,m}\left(1 - \frac{P_m}{K_m}\right)$$

2.3.4 Model for International Migrant Stock

Based on the assumption in section 2.1.3, the world population is proportionate to the annual migrant stock. According to this, we can construct a simple function to represent this relationship: $M = aP+C$, where a is a coefficient and C is a constant for all the other factors (such as environment, politics-see section 2.1.3). However, we found that this model is too simple and unfit for the purpose of predicting

the migration trends. Therefore, to complicate the model, we decided to add more features using polynomial regression and came up with a more exhaustive function:

$$M = \theta_0 + P\theta_1 + P^2\theta_2 + ... + P^n\theta_n$$

where θ_i are coefficients.
Taking derivative of M with respect to t, we get the differential equation to apply in the main model:

$$\frac{dM}{dt} = \frac{dP}{dt}(\theta_1 + 2P\theta_2 + ... + nP^{n-1}\theta_n)$$

2.3.5 Model for Language Utility

In this model, we consider three important indices to reflect the utility of each language in terms of its economic potential: GDP, GDP growth rate, and GDP per capita. Using the statistics of all the countries that speak the current top fourteen languages, we scale the data of each index, with the greatest one being 100 and the smallest 1. Then, we create a point-based system to weight the three indices:

1. GDP per capita (G_c): 5 points
2. GDP growth rate (G_r): 1 points
3. GDP (G_t): 4 points[3]

These statistics are calculated for all 14 languages based on all the countries that use the language. We find the average of the GDP per capita, the growth rate, and the sum of their cumulative GDPs.
For purposes of calculations, we only consider 4 years: 1970, 1996, 2012, and 2020. These years are representative of two 60-year periods: from 1950 to 2010 and 2010 to 2070. We find the mean of the values for both pair of years: 1970 and 1996, 2012 and 2020.
Let $A_{14\times3}$ be a matrix of all the means of the economic statistics for the period 1970 - 2010. The first column is the GDP per capita; the second is the growth

[3]This system is completely based on our subjective opinion about the importance of each economic index and is subjected to change as necessary. For future development, more indices can be added to complicate the model. These indices can also be weighted differently.

rate, and the last is the GDP. Each row of the matrix is for one of the fourteen languages considered.

$$A = \begin{bmatrix} G_{c_1} & G_{r_1} & G_{t_1} \\ G_{c_2} & G_{r_2} & G_{t_2} \\ \vdots & \ddots & \vdots \\ G_{c_14} & G_{r_14} & G_{t_14} \end{bmatrix}$$

Let $\vec{v}$ be the weight of the three indices: $\vec{v} = \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}$. We can calculate the weighted indices by finding the matrix multiplication of A and $\vec{v}$: $\begin{bmatrix} G_{c_1} & G_{r_1} & G_{t_1} \\ G_{c_2} & G_{r_2} & G_{t_2} \\ \vdots & \ddots & \vdots \\ G_{c_14} & G_{r_14} & G_{t_14} \end{bmatrix} \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}$

The result will be $[G_i]_{14\times 1}$, where G_i would be the weighted economic index of the i^{th} languages. We then find the total of all these indices and let the sum be S. The language utility of the i^{th} language is:

$$U_i = \frac{G_i}{S}$$

This will give us a general estimate of the importance of the language in the global market where these fourteen languages are the predominant ones. We then use this index in our main model to project the change in the distribution of languages, particularly to predict the number of new L2 speakers who will adopt the language for its economic usability.

According to our result, English would be at the top of the language utility ranking, with the score of 20.53%. Next is German, with countries speaking the language in the top of most global economic rankings. Mandarin scores the third position with its fast-growing economy. Malay surprises us with its performance, outranking French and Spanish to be the 7^{th} most potential language in terms of economic growth. The rest of the languages in the table are relatively predictable.

2.3.6 Model Testing

Applying the world population data from 1950 to 2010 to the logistic growth model, using numerical solving for for differential equation, we found that $k = 0.025445$ and $K = 13539.1159$.

To find the coefficients θ_i in the International Migrant Stock Model, we use migration data from 1960 to 2015, with a 5-year interval.[6]

$$M_i = \theta_0 + P_i\theta_2 + ... + {P_i}^n\theta_n, i \in [1960, 2015]$$

$$\text{Let } \theta = \begin{bmatrix} \theta_0 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}, X = \begin{bmatrix} 1 & P_{1960} & \dots & {P_{1960}}^n \\ 1 & P_{1965} & \dots & {P_{1965}}^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & P_{2015} & \dots & {P_{2015}}^n \end{bmatrix}, y = \begin{bmatrix} M_{1960} \\ M_{1965} \\ \vdots \\ M_{2015} \end{bmatrix}$$

Using normal equation: $\theta = (X^T X)^{-1} X^T y$ [9]
We get that, for $3 <= i <= n$, θ_i is too insignificant. Therefore, we only consider n = 3 with coefficients $\theta_0 = 65.400671$, $\theta_1 = -0.012232$, and $\theta_2 = 0.00005$
We get the model of International Migrant Stock:

$$M = 65.400671 - 0.012232 \times P + 0.00005 \times P^2$$

$$\frac{dM}{dt} = \frac{dP}{dt}(-0.012232 + 0.0001 \times P)$$

We have calculated the language utilities for the period of 1950 - 2010 (Figure 1). This information will be used to predict the distribution of languages in the year 2015. Plugging these values into the main model, we can find the total number of speakers of each language. Dividing this number by the world population, we get the percentage of the language speakers (language distribution). When

#Rank	Languages	1950-2010
1	English	23.63
2	German	18.16
3	French	8.91
4	Japanese	7.50
5	Spanish	6.33
6	Russian	6.10
7	Portuguese	4.83
8	Arabic	4.48
9	Malay	4.45
10	Mandarin	4.29
11	Hindi	3.53
12	Hausa	2.82
13	Punjabi	2.55
14	Bengali	2.40

Figure 1: Languages Utilities in the Period between 1950 and 2010 (%)

comparing these results with the 2015 Ethnologue reports, we found some similarities and inconsistencies (Figure 2 & 3). Our model performs quite accurately with English, Spanish, Bengali, Portuguese, Russian, Japanese, and German. Results for languages like Mandarin, Hindi, Arabic, French, and Punjabi are off by 100-400 million speakers. This error is because of the assumption that all the people of a country would speak its official language. Some countries have several official/major languages, and our method cannot accurately model the language distribution within each country for such a large set of data. Therefore, this model would perform relatively well for countries where most of the population use the same language. Countries like India and China are too complicated to model and require more variables and data.

# Rank	Language	L1 Native Speaker (million)	Total (million)
1	Mandarin	848	1026
2	Spanish	399	488.5
3	English	335	840
4	Hindi	260	380.5
5	Arabic	242	279
6	Portuguese	203	203
7	Bengali	189	208.2
8	Russian	166	166
9	Japanese	128	128
10	Malay	124.5	127.5
11	German	78.1	91.3
12	Frances	75	162
13	Punjabi	62.6	62.6

Figure 2: Real data

#Rank	Languages	Total (Million)	Native Speaker L1 (Million)
1	Mandarin	1,453.30	897
2	English	812.80	371
3	Hindi	719.68	329
4	Arabic	479.89	290
5	Spanish	454.88	436
6	French	309.31	76
7	Malay	294.74	90
8	Bengali	284.81	242
9	Portuguese	277.79	218
10	Punjabi	230.82	148
11	Russian	184.13	153
12	Japanese	131.84	128
13	Hausa	123.98	85
14	German	119.71	76

Figure 3: Model data

3 Language Distribution Projections

3.1 Number of Language Speakers

Using the same method as the Model Testing section (see section 2.3.6), we calculated the population size of the most spoken languages in the world in the year 2070.

# Rank	Language	Population of each languages 2070 (million)
1	Mandarin	1,718.21
2	English	1,621.76
3	Hindi	1,417.32
4	Arabic	1,195.77
5	Spanish	628.85
6	French	756.73
7	Malay	420.67
8	Bengali	492.05
9	Potuguese	406.39
10	Punjabi	657.43
11	Russian	185.32
12	Japanese	139.18
13	Hausa	456.25
14	German	167.25

Figure 4: Number of Speakers of the Most Spoken Languages in 2070

3.2 Top-Ten Language List (Figure 5)

The first three ranks remain the same as in 2017. Arabic climbs one position. French jumps 5 ranks to be the fifth most spoken language. Punjabi makes its debut into the top ten with an impressive jump of 6^{th} positions. Hausa and Portuguese swap places, with Hausa as the ninth most spoken language. Bengali and Portuguese remain the same. Russian drops 5 ranks to position 12.

The jump of French is expectable because many African francophone countries are rapidly developing. They also have high population growth rates. This explains the growth in the number of French speakers. Punjabi's debut in the top ten can be accounted by India's fast-growing economy and population. Hausa's appearance in the list is caused by Nigeria's rapid population growth. Meanwhile, the Russian Federation's economy has been slowing down, and its population also remains much stable for the long term.

In terms of percentage, most languages show an increase in their portions in the world population, except for Spanish and Russian. A dramatic increase can be observed with Hindi, Arabic, French, Punjabi, and Hausa. These are mostly developing countries with an expanding population and economy. Other more developed countries are slowing down in terms of population and economic growth, accounting for the drop of Spannish, Russian, and Portuguese. Mandarin and English remain the most important languages in the world, making up about 30% of the world population.

#Rank	Year / Language	2017	2070	Ranking Movement
1	Mandarin	14.34%	15.30%	+0
2	English	12.93%	14.44%	+0
3	Hindi	7.16%	12.62%	+0
4	Arabic	5.55%	10.65%	+1
5	French	3.01%	6.74%	+5
6	Punjabi	1.95%	5.86%	+6
7	Spanish	6.93%	5.60%	-1
8	Bengali	3.43%	4.38%	+0
9	Hausa	1.97%	4.06%	+2
10	Malay	3.70%	3.75%	+4
11	Portuguese	3.01%	3.62%	-2
12	Russian	3.51%	1.65%	-5
13	German	1.70%	1.49%	+1
14	Japanese	1.70%	1.24%	-1

Figure 5: Percentage of Language Speakers in the World Population in 2070

3.3 Geographic Distribution

In reality, the language utility index that we proposed above only model the dynamics of the languages rather than the dynamics of migrants speaking them. In other words, this index is more indicative of the shifts in the number of speakers in general, rather than their geographic locations. However, we can still make some predictions based on the percentage difference between the number of language speakers in 2015 and 2070 (figure 6).

We observe that from 2015 to 2070, the difference in percentage of native speakers only accounts for less than half of the difference with total number of speakers. The remaining is because of changes in L2. According to our assumptions, the only variable affecting the number of L2 speakers is the language utility index. The index signifies the potential of migrants to a particular language's area. This means that there are changes to the geographic distribution of language speakers around the world. Changes in the geographic distribution of the world's major languages mostly depend on future trends in international migration.

Language	Total speaker(%)			Native spearker(%)		
	2070	2015	Difference (%)	2070	2015	Difference (%)
English	14.44	11.01	3.43	6.28	5.03	1.25
Mandarin	15.30	19.69	-4.39	9.22	12.15	-2.93
Hindi	12.62	9.75	2.87	5.72	4.46	1.26
Arabic	10.65	6.50	4.15	6.33	3.93	2.40
Russian	1.65	2.49	-0.84	1.28	2.07	-0.79
German	1.49	1.62	-0.13	0.66	1.03	-0.37
Japanese	1.24	1.79	-0.55	0.96	1.73	-0.77
Punjabi	5.86	3.13	2.73	3.66	2.01	1.66
Portuguese	3.62	3.76	-0.14	2.73	2.95	-0.22
Hausa	4.06	1.68	2.38	2.72	1.15	1.57
Spanish	5.60	6.16	-0.56	5.23	5.91	-0.68
Bengali	4.38	3.86	0.52	3.62	3.28	0.34
Malay	3.75	3.99	-0.25	1.07	1.22	-0.15
French	6.74	4.19	2.55	1.59	1.03	0.56
Total	91.41	79.64	11.77	51.06	47.95	3.11

Figure 6: Percentage difference between 2070 and 2015

4 Reports on Opening International Offices

4.1 Locating the International Offices

According to our model, the most important index that represents the economic potential of each language is the language utility. Locating the offices where the local language has commercial potential in the global market would be advisable. English is already a global language. Therefore, offices should be located where there is a supply of local labors capable of using both English and a potential local language with high utility. Fast-growing economies like India where people can use a variety of important languages like English, Hindu, and Punjabi are desirable places for the international offices.
Based on our predictions we made about the top ten languages in 2070, we will locate 6 offices in these countries:

1. India: India is our top choice because it can speak a variety of important languages like English, Hindi, Bengali, and Punjabi. India also has the second largest population in the world. Its economy has a lot of potential to expand in the near future. Languages used in the office: English, Hindi, Bengali, Punjabi.

2. Saudi Arabia: This is a largest country in the Middle East, an important economic zone. This dynamic area has a lot of potential for development. Saudi Arabia has one of the biggest economy in the area, with a first-ranking

#Rank	Languages	2010-2070
1	English	17.93
2	Mandarin	12.52
3	Hindi	3.15
4	Arabic	4.31
5	Russian	4.19
6	German	16.73
7	Japanese	9.80
8	Punjabi	3.11
9	Portuguese	4.29
10	Hausa	1.72
11	Spanish	4.63
12	Bengali	3.53
13	Malay	7.84
14	French	6.25

Figure 7: Language Utilities for period 2015 to 2070 (%)

GDP.[7] Saudi Arabia is a good choice for a stepping stone into the Arab-speaking world. Languages used in the office: English, Arabic.

3. Brazil: As a representative of South America, Brazil has a dynamic economy dominated by Portuguese native speakers, an important language in 2070. It has also had an impressive performance on economic rankings based on a variety of indices. Brazil's main language, Portuguese, is the eleventh language in the list. Languages used in the office: English, Portuguese.

4. France: The only European representative in the list, France is our top choice for francophone countries. With a population of 67.15 million people (October 2017), France dominates as a major economy in Europe that has always been leading the world in the last centuries.[20] Access to France would also allow for access into the European Union's economy, especially after Brexit. Languages used in the office: English, French.

5. Malaysia: Malaysia is an important country in Asia. In addition to using Malay, the tenth most spoken language, Malaysia also plays a critical role in the economic dynamics of South East Asia. This is an essential area for future investments because of its critical location and major source of labor. Languages used in the office: English, Malay.

6. Nigeria: As the only African country in the list, Nigeria plays an important role in extending our country's reach to the realms of Africa. Nigeria is also one of the few countries in the world speaking Hausa, a language projected

to take a large portion of the world's population. The number of people speaking Hausa is estimated to continue to expand in the future. Languages used in the office: English, Hausa.

The only languages not featured in the top ten is English and Mandarin, whose representatives are already locations for the company's current offices. All of these offices will feature employees capable of the most important languages in the world (from rank 1 to rank 11).

4.2 Long-term Analysis

Because our model depends greatly on population growth rates and global economies, predictions can only be made with sufficient data of these indices. Projections in the long run will require estimated data further into the future. Depending on the current economic outlook, prospective economies like China and India are still expanding rapidly. [5] Languages like Hindi and Chinese Mandarin would attract more prospective speakers at least until these countries' economic development rates slow down significantly. In addition, English-speaking countries like the United States, the United Kingdom, and Australia, which have for long been major destinations for international migrants, would still maintain these trends, unless their governments tighten immigration policies, which is happening to America[8]. However, this would not change the reality that there are already many immigrants of other languages in these countries. The language distribution within these countries would still remain relatively diverse. As a result, there would not be many changes in the geographic distribution of languages in the long run, should future global economic growth and immigration policies remain the same. The only major exception to this would be in the case of political instability, which would constitute a powerful force to alter the trends of migration and affect the world's distribution of language speakers.

4.3 Further Suggestions

We would not suggest opening less offices, because all of the chosen locations represent different regions in the world with great economic potential and diversity. However, for purposes of saving resources, there are several options for opening less than 6 offices. 4 of the current 8 offices are in Asia. If necessary, we can relocate the offices to multilingual countries like Australia. This can also be done with offices in the Middle East, Africa, and South America. The best option would be

to locate the offices in countries that have a number of important languages. The company should choose multilingual regions (Singapore), or regions where there are a lot multinational immigrants (Australia). This way, we can substitute two ore more monolingual locations in our list with one multilingual location. Doing this would require further research into the internal distribution of languages in these multilingual countries and their economic prospects. Additional information like geographic and political attributes must be taken into account when considering this option. In the best scenario, we would like to build a model that rates the potential of each country in terms of the languages they use and their economic power. The model would be customized based on the requirements and interests of the company. Other factors of interests can also be included such as the labor force, minimum wage, and other values.

5 Limitations & Future Improvement

5.1 Limitations

The implementation of our model is closely based on our initial assumptions. There are several limitations in our assumptions that we would like to have been able to address:

1. In this model, we are only concerned with a limited number of factors that influence the shifts in the distribution of language speakers around the world (*de jure* languages, population growth, migration trends, economical factors). There are other influences that have an impact on the number of people speaking a specific language, such as lexical difficulty, *de facto* status, distribution of language speakers within a country, second & third language speakers, cultural assimilation of different groups, tourism, technology, etc.

2. As mentioned above, it is not necessary (or accurate) that the entire population of a country knows its official language(s). This assumption has thrown our results off by a few hundred millions because of its generality and inaccuracy. Another similar problem applies to the L2 speakers assumption. While we speculate that all of the immigrants to a country would be able to speak its native language(s), this is not necessarily true in many cases. As we have identified, further research can be done to classify the distribution of languages within each country to fine-tune our model.

3. Language utility, an important concept in our model, is very much subjective to our opinion of what would be the best indices for the economic growth of a country. There are many other indices that could be taken into account (unemployment rate, inflation rate, etc.). In addition, this is a very narrow method to find the utility of a language, because it cannot be based solely on the economy. Various other factors affect the utility of a language, as mentioned above (e.g. lexical difficulty, population density, political stability, environment). Furthermore, how we weighted the three economic indices (GDP, GDP per capita, GDP growth rate) is, once again, very biased. The accuracy of this model is based on careful analysis and categorization of relevant factors. The more complicated the factors are and the better we weight them, the resulting index will be more indicative of the actual usability of a language.

4. The years we chose for the purpose of estimating the language utility also are not very representative of the entire period. We used 1970 and 1996 for the period between 1950 and 2010, and 2012 and 2020 for the time period 2010 and 2070. The second period is especially skewed because we did not have enough information about the world economy a long time into the future.

5. The 14 languages chosen for analysis are also subject to change. Here, we only consider the top languages with the most number of speakers. However, the 15^{th} position on the 2017 rank, Persian, has only 8 million speakers less than the 14^{th} one, German. It is, nevertheless, not chosen. We did not include more languages because there was not enough time and people to process such large data sets. The 2070 list might have appeared differently with the inclusion of other languages, but we estimate that these discrepancies would not be too significant.

6. The 6 locations that we proposed for the opening of new offices (see section 4.1) are mainly handpicked based on general knowledge. We would like to have a more systematic model standardized around certain indices valued by the company.

5.2 Future Improvement

As explained in the limitations section (see section 5.1), our proposed model for the distribution of language speakers has a lot of room for improvement. Further

research and work on this could be carried out in a number of directions based on the original model. The language utility index could be modified to allow for more factors indicative of the changes in the number of speakers. Other influences on language distribution should be considered to complicate the model. We would also like to do some further testing to verify the proportional relationship between population growth rate and the growth rate of the population speaking a specific language. Further investigation of in-country language distribution and their dynamics would give us more insight into L1 and L2 speakers within a country. Finally, we can expand this model to consider all the living languages in the world and predict the future trends of language distribution on a more complete scale. The proposed list of locations for new offices could use more research and revision, using a customized model to rate different regions of the world.

Building this model, we are aware of its shortcomings and inconsistencies. However, we believe that these limitations can be addressed with more comprehensive data and research, and more effective methods to rate the usability of a language.

6 Memorandum

TO: Chief Operating Officer
FROM: MCM Consulting Team
DATE: February 12, 2018
SUBJECT: Reports on Global Languages Trends and Proposal for Office Locations

Context & Objectives

The world's demography has undergone significant changes in the past few decades. This has a direct impact on the use of languages around the world, making it imperative for companies and businesses to anticipate the affects of global language dynamics on their business ventures. Understanding your needs and concerns, we have come up with a method to predict the future population size of a language based on the natural population growth rates, global migration trends, and other economic factors. Having made an estimate of the distribution of language speakers in 2070, we then decide on six potential locations for your new offices that will ensure the multinational and multilingual values of your company.

Method

First, we picked the 14 most spoken languages to consider: Mandarin Chinese, English, Hindi, Spanish, Arabic, Malay, Russian, Bengali, Portuguese, French, Hausa, Punjabi, Japanese, and German. Then we found all of the countries whose official language(s) belong in these groups, and categorized them accordingly. After we had the total population of each language group, we used the logistic growth model to decide the growth rate of each population size. We speculated that the number of speakers for each language would be proportional to its population size.
In addition, we consider another index called language utility. This is used to measure (in percentage) the share of a language in the global market. This measurement is indicative of the usability of a language in terms of its economic power. To calculate this number, we consider three factors: GDP, GDP growth rate, and GDP per capita. We also find the cumulative/average values of these indices for each language group, and weighted them according to their importance. The indices of GDP, GDP growth rate, and GDP per capita are weighted 4, 1, 5 respectively.

Analysis Results

According to this model, Mandarin, English, and Hindi remained the three most important languages in 2070. The order of the other languages in the top 10 most spoken ones is relatively similar to their current order, except for French which made an impressive leap from the 10^{th} to the 5^{th} place. Punjabi makes its first debut in the top ten list, climbing 6 ranks to the 6^{th} position. Portuguese and Russian drop out of the top ten languages. Hausa replaces Portuguese as the 9^{th} most spoken language. Malay drops to the 10^{th} position.

According to this ranking, we then decided on the locations for the new offices. We propose the following 6 locations, which satisfy your requirement about languages and meet our demands regarding their economic potential: India, Saudi Arabia, Brazil, France, Malaysia, and Nigeria. Employees at these locations are capable of the first 11 most spoken languages in the world. Some locations, like India, have a multilingual background of up to 3-4 languages. These locations are picked in addition to your current offices in Shanghai and New York.

Alternatively, your company can opt for less offices. In this case, we will change the current locations to countries which use several major languages, so that one office can represent several different languages. However, this approach would require more investigation into the in-country language distribution and other indices to rate each country's economic potential. You are by no means required to use our suggested list. What we propose is just an example of the combinations you can make to ensure that the offices are well spread out and well integrated on a global scale.

We are glad to further discuss these recommendations with you and follow through on your future decisions. You are welcome to make any modifications to this proposal as fit.

Best,
MCM Consultant Team

7 Conclusion

As the dynamics of the world's demography are constantly shifting, the need arises for a method to predict the changes in its make-up. In this paper, we proposed a way to model the language attribute of the world's demography, using its population growth rate, projected migration trends, and economic indices. This model allows us to investigate the changes of each language's population over time both in terms of L1 and the total number of speakers. These changes reflect the population growth of the language and the movement of speakers from one to more languages as indicated by migration trends.

The proposed language utility index is also an important part of our model. In considering indices such as GDP, GDP growth rate, and GDP per capita, we tried to predict the trends of migrants to regions where their languages are more valuable in terms of economic power. We speculated that the number of speakers of a language correlates with that language's economic utility. This is why we introduce this index in the model, in addition to the population growth rate.

That said, our model is not without certain limitations. Our biggest concern was that we dealt mainly with initial assumptions. This led to certain unavoidable pitfalls in our results. In addition, we only considered a limited number of languages and factors of influence, which allowed us to simplify the model. This limitation can be addressed by scaling the model and adding more variables to complicate the model.

Using the results generated by the model, we handpicked 6 locations for building offices that satisfy the multilingual requirements. We also considered the economic aspect of these locations to ensure that they would be suitable for future endeavors. The locations are well dispersed and well integrated with the local region to allow for access into the regional economies. Further research can be done to streamline the process of choosing these locations. With more data, we can rate each country according to the development indices and their language utility to decide the best places to open more offices.

Although we only considered a small number of countries, languages, and factors, this model can be scaled to include all the countries and languages in the world to model the actual reality more comprehensively. Other factors affecting the movement of languages and people should be considered for further improvement. Economic indices are our main reference, but we can also include environmental, political, and social factors in the language utility index.

References

[1] Summary by world area. (n.d.). Retrieved February 11, 2018, from https://www.ethnologue.com/statistics

[2] "Constitutional Provisions Official Language Related Part-17 Of The Constitution Of India". National Informatics Centre (in Hindi). Archived from the original on 8 November 2016. Retrieved 1 December 2017.

[3] "World Population Prospects: The 2017 Revision". ESA.UN.org (custom data acquired via website). United Nations Department of Economic and Social Affairs, Population Division. Retrieved 10 September 2017.

[4] The International Migration Report 2017 (Highlights) — Multimedia Library - United Nations Department of Economic and Social Affairs. (n.d.). Retrieved February 11, 2018, from https://www.un.org/development/desa/publications/international-migration-report-2017.html

[5] World Economic Outlook (October 2017) - Real GDP growth. (n.d.). Retrieved February 11, 2018, from http://www.imf.org/external/datamapper/NGDP_RPCH@WEO

[6] United Nations Population Division — Department of Economic and Social Affairs. (n.d.). Retrieved February 12, 2018, from http://www.un.org/en/development/desa/population/migration/data/estimates2/estimates17.shtml

[7] GDP (current US$). (n.d.). Retrieved February 12, 2018, from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD The World Bank.

[8] Bennett, B. (n.d.). Trump pushes to sharply cut the number of legal immigrants and move U.S. to a merit-based immigration system. Retrieved February 12, 2018, from goo.gl/KW7RKu

[9] Derivation of the Normal Equation for linear regression - Eli Benderskys website. (n.d.). Retrieved February 12, 2018, from https://eli.thegreenplace.net/2014/derivation-of-the-normal-equation-for-linear-regression

[10] GDP Annual Growth Rate - Forecast - 2018-2020. (n.d.). Retrieved February 12, 2018, from https://tradingeconomics.com/forecast/gdp-annual-growth-rate

[11] International Migrants by Country of Destination, 1960-2015. (2014, February 7). Retrieved February 12, 2018, from https://www.migrationpolicy.org/programs/data-hub/charts/international-migrants-country-destination-1960-2015

[12] Logistic Model. (n.d.). Retrieved February 12, 2018, from https://www.ma.utexas.edu/users/davis/375/popecol/lec5/logist.html

[13] PricewaterhouseCoopers. (n.d.). The World in 2050. Retrieved February 12, 2018, from https://www.pwc.com/gx/en/issues/economy/the-world-in-2050.html

[14] Snapshot. (n.d.). Retrieved from http://www.ethnologue.com/18/

[15] Stockton, N., Stockton, N. (n.d.). Where everyone in the world is migratingin one gorgeous chart. Retrieved February 12, 2018, from https://qz.com/192440/where-everyone-in-the-world-is-migrating-in-one-gorgeous-chart/

[16] These are the most powerful languages in the world. (n.d.). Retrieved February 12, 2018, from https://www.weforum.org/agenda/2016/12/these-are-the-most-powerful-languages-in-the-world/

[17] UNdata — record view — Net number of migrants, both sexes combined (thousands). (n.d.). Retrieved February 12, 2018, from http://data.un.org/Data.aspx?q=migrant&d=PopDiv&f=variableID%3a84

[18] World. (n.d.). Retrieved February 12, 2018, from http://www.ethnologue.com/18/

[19] International migrant stock, total — Data. (n.d.). Retrieved February 12, 2018, from https://data.worldbank.org/indicator/SM.POP.TOTL

[20] "Field Listing :: Area". The World Factbook. CIA. Retrieved 1 November 2015.

Appendix A: Collected Data

Language	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
English	316 129	320 198	324 758	329 708	334 963	340 448	346 102	351 877	357 736	363 651	369 600	375 561	381 506	387 401	393 208	398 908	404 480	409 943	415 355	420 799
Mandarin	563 065	578 572	591 313	602 186	611 936	621 162	630 302	639 647	649 353	659 469	670 021	681 097	692 926	705 912	720 537	737 090	755 733	776 222	797 919	819 928
Hindi	201 999	205 180	208 562	212 120	215 838	219 703	223 712	227 868	232 178	236 657	241 315	246 160	251 191	256 391	261 742	267 232	272 851	278 613	284 561	290 756
Arabic	85 269	87 371	89 570	91 861	94 239	96 705	99 258	101 903	104 644	107 486	110 435	113 495	116 665	119 942	123 319	126 792	130 371	134 061	137 859	141 756
Rusian	120 198	121 852	123 775	125 886	128 120	130 419	132 743	135 060	137 351	139 605	141 814	143 966	146 038	147 995	149 799	151 423	152 852	154 102	155 226	156 298
German	90 518	90 949	91 381	91 811	92 239	92 669	93 110	93 574	94 080	94 644	95 280	95 992	96 771	97 586	98 401	99 184	99 926	100 624	101 259	101 807
Japanese	82 802	84 316	85 659	86 869	87 980	89 018	90 004	90 954	91 878	92 782	93 674	94 561	95 459	96 390	97 380	98 447	99 595	100 820	102 120	103 492
punjabi	48041.8491	48639.925	49324.1185	50088.8405	50929.592	51842.9026	52826.4287	53878.8579	54999.8314	56189.6897	57448.802	58776.9015	60172.4468	61632.3454	63152.7225	64731.117	66368.813	68067.9538	69828.4851	71650.4256
Portugese	74 721	76 619	78 593	80 619	82 684	84 780	86 906	89 071	91 287	93 571	95 939	98 398	100 939	103 540	106 167	108 796	111 421	114 051	116 704	119 407
Hausa	25 276	25 693	26 138	26 607	27 097	27 607	28 135	28 683	29 253	29 848	30 471	31 124	31 806	32 513	33 242	33 990	34 758	35 549	36 367	37 216
Spanish	135 947	138 915	142 012	145 223	148 537	151 947	155 456	159 069	162 800	166 662	170 667	174 822	179 117	183 530	188 030	192 594	197 207	201 876	206 616	211 454
Bengali	71 350	72 687	74 032	75 430	76 914	78 506	80 217	82 049	83 994	86 036	88 159	90 353	92 622	94 985	97 466	100 080	102 849	105 750	108 694	111 557
Malay	76 723	78 260	79 934	81 741	83 677	85 736	87 913	90 201	92 593	95 083	97 664	100 332	103 082	105 913	108 824	111 816	114 884	118 025	121 234	124 507
French	104148.098	105318.647	106612.078	108001.236	109466.31	110994.387	112580.318	114226.073	115939.642	117733.569	119618.912	121598.764	123663.059	125785.175	127928.107	130066.102	132184.455	134286.709	136387.865	138515.506

Language	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
English	426 340	432 000	437 774	443 672	449 694	455 843	462 131	468 566	475 140	481 839	488 651	495 574	502 612	509 784	517 110	524 605	532 284	540 138	548 132	556 216
Mandarin	841 554	862 631	883 192	902 987	921 781	939 438	955 836	971 069	985 506	999 676	1 014 006	1 028 470	1 043 070	1 058 288	1 074 717	1 092 699	1 112 499	1 133 795	1 155 613	1 176 602
Hindi	297 239	304 028	311 107	318 434	325 944	333 594	341 361	349 258	357 319	365 592	374 111	382 882	391 877	401 050	410 336	419 685	429 080	438 523	448 009	457 541
Arabic	145 754	149 841	154 038	158 406	163 031	167 969	173 245	178 835	184 685	190 713	196 857	203 098	209 445	215 900	222 477	229 182	236 003	242 924	249 941	257 053
Rusian	157 376	158 480	159 603	160 745	161 895	163 048	164 204	165 375	166 571	167 805	169 086	170 405	171 752	173 128	174 535	175 964	177 422	178 882	180 262	181 452
German	102 252	102 586	102 813	102 946	103 006	103 009	102 966	102 880	102 761	102 622	102 473	102 318	102 170	102 060	102 028	102 099	102 282	102 566	102 946	103 408
Japanese	104 926	106 425	107 977	109 533	111 030	112 423	113 690	114 839	115 890	116 878	117 827	118 743	119 616	120 438	121 200	121 894	122 520	123 083	123 595	124 069
punjabi	73535.5995	75484.8123	77503.8422	79605.825	81807.5351	84122.219	86551.2514	89094.7633	91762.4507	94565.2329	97508.4041	100592.938	103809.181	107137.519	110551.39	114027.988	117561.622	121145.99	124757.087	128366.824
Portugese	122 181	125 028	127 947	130 939	134 011	137 163	140 390	143 689	147 058	150 499	154 006	157 591	161 239	164 888	168 454	171 883	175 141	178 258	181 317	184 441
Hausa	38 099	39 015	39 966	40 968	42 045	43 209	44 471	45 820	47 228	48 653	50 065	51 454	52 828	54 207	55 619	57 084	58 608	60 185	61 810	63 476
Spanish	216 407	221 478	226 650	231 897	237 179	242 470	247 758	253 047	258 330	263 609	268 881	274 144	279 396	284 645	289 903	295 176	300 463	305 760	311 085	316 458
Bengali	114 261	116 762	119 107	121 414	123 851	126 540	129 519	132 754	136 197	139 770	143 415	147 117	150 893	154 742	158 674	162 690	166 790	170 957	175 161	179 366
Malay	127 841	131 231	134 676	138 173	141 718	145 308	148 937	152 602	156 312	160 074	163 894	167 773	171 701	175 655	179 609	183 541	187 442	191 311	195 150	198 961
French	140691.472	142929.674	145223.264	147552.428	149887.086	152207.153	154506.31	156808.386	159158.141	161612.471	164213.496	166969.576	169868.692	172904.805	176069.445	179350.094	182763.378	186319.183	189990.192	193724.286

Language	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
English	564 358	572 537	580 772	589 127	597 693	606 529	615 672	625 095	634 710	644 397	654 073	663 688	673 283	682 975	692 929	703 258	714 010	725 123	736 458	747 819
Mandarin	1 195 770	1 212 799	1 227 910	1 241 320	1 253 451	1 264 647	1 274 912	1 284 200	1 292 768	1 300 937	1 308 953	1 316 942	1 324 921	1 332 900	1 340 840	1 348 717	1 356 546	1 364 366	1 372 195	1 380 051
Hindi	467 120	476 733	486 372	496 050	505 787	515 594	525 471	535 402	545 362	555 320	565 247	575 126	584 952	594 720	604 427	614 069	623 647	633 144	642 514	651 701
Arabic	264 253	271 555	278 944	286 357	293 706	300 938	308 038	315 048	322 049	329 155	336 457	343 967	351 693	359 714	368 122	376 977	386 313	396 097	406 227	416 560
Rusian	182 378	183 011	183 373	183 497	183 437	183 240	182 920	182 484	181 964	181 400	180 824	180 254	179 708	179 225	178 849	178 611	178 526	178 588	178 790	179 112
German	103 934	104 533	105 191	105 849	106 428	106 875	107 159	107 300	107 351	107 390	107 476	107 629	107 833	108 055	108 246	108 375	108 427	108 424	108 415	108 465
Japanese	124 516	124 940	125 341	125 718	126 063	126 375	126 654	126 903	127 127	127 336	127 534	127 724	127 903	128 068	128 214	128 336	128 433	128 505	128 551	128 567
punjabi	131955.338	135507.151	139025.124	142530.773	146057.055	149626.618	153249.46	156914.595	160598.94	164269.228	167903.405	171495.669	175059.761	178617.378	182199.488	185830.578	189513.171	193246.081	197045.002	200927.717
Portugese	187 715	191 171	194 777	198 482	202 205	205 889	209 516	213 106	216 676	220 261	223 882	227 545	231 237	234 935	238 608	242 236	245 809	249 336	252 838	256 342
Hausa	65 175	66 909	68 682	70 499	72 365	74 284	76 260	78 294	80 388	82 544	84 764	87 050	89 406	91 840	94 363	96 982	99 701	102 519	105 437	108 455
Spanish	321 890	327 394	332 955	338 528	344 049	349 474	354 788	360 008	365 168	370 317	375 493	380 692	385 908	391 171	396 519	401 970	407 547	413 224	418 915	424 500
Bengali	183 544	187 676	191 767	195 844	199 944	204 094	208 293	212 524	216 768	221 000	225 197	229 362	233 485	237 523	241 421	245 143	248 668	252 013	255 232	258 403
Malay	202 747	206 508	210 242	213 944	217 607	221 229	224 809	228 353	231 880	235 414	238 973	242 564	246 185	249 837	253 518	257 228	260 964	264 722	268 501	272 295
French	197487.807	201246.738	204994.767	208752.816	212581.246	216534.797	220651.768	224931.094	229364.371	233916.287	238554.176	243261.82	248065.462	253008.371	258150.012	263533.896	269177.367	275061.275	281150.376	287395.046

Language	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
English	625 095	634 710	644 397	654 073	663 688	673 283	682 975	692 929	703 258	714 010	725 123	736 458	747 819	759 065	770 146	781 109	792 022	802 995	814 109
Mandarin	1 284 200	1 292 768	1 300 937	1 308 953	1 316 942	1 324 921	1 332 900	1 340 840	1 348 717	1 356 546	1 364 366	1 372 195	1 380 051	1 387 932	1 395 841	1 403 733	1 411 494	1 418 972	1 426 050
Hindi	535 402	545 362	555 320	565 247	575 126	584 952	594 720	604 427	614 069	623 647	633 144	642 514	651 701	660 666	669 386	677 877	686 189	694 394	702545.082
Arabic	315 048	322 049	329 155	336 457	343 967	351 693	359 714	368 122	376 977	386 313	396 097	406 227	416 560	426 983	437 469	448 024	458 613	469 203	479 774
Rusian	182 484	181 964	181 400	180 824	180 254	179 708	179 225	178 849	178 611	178 526	178 588	178 790	179 112	179 535	180 061	180 675	181 333	181 975	182 554
German	107 300	107 351	107 390	107 476	107 629	107 833	108 055	108 246	108 375	108 427	108 424	108 415	108 465	108 619	108 894	109 268	109 709	110 165	110 598
Japanese	126 903	127 127	127 336	127 534	127 724	127 903	128 068	128 214	128 336	128 433	128 505	128 551	128 567	128 552	128 505	128 426	128 313	128 163	127 975
punjabi	156914.595	160598.94	164269.228	167903.405	171495.669	175059.761	178617.378	182199.488	185830.578	189513.171	193246.081	197045.002	200927.717	204904.543	208982.15	213151.07	217384.481	221644.931	225903.119
Portugese	213 106	216 676	220 261	223 882	227 545	231 237	234 935	238 608	242 236	245 809	249 336	252 838	256 342	259 870	263 428	267 008	270 602	274 192	277 767
Hausa	78 294	80 388	82 544	84 764	87 050	89 406	91 840	94 363	96 982	99 701	102 519	105 437	108 455	111 573	114 791	118 110	121 524	125 025	128 606
Spanish	360 008	365 168	370 317	375 493	380 692	385 908	391 171	396 519	401 970	407 547	413 224	418 915	424 500	429 897	435 067	440 036	444 860	449 625	454 393
Bengali	212 524	216 768	221 000	225 197	229 362	233 485	237 523	241 421	245 143	248 668	252 013	255 232	258 403	261 583	264 791	268 014	271 235	274 429	277575.785
Malay	228 353	231 880	235 414	238 973	242 564	246 185	249 837	253 518	257 228	260 964	264 722	268 501	272 295	276 099	279 913	283 724	287 506	291 219	294 838
French	224931.094	229364.371	233916.287	238554.176	243261.82	248065.462	253008.371	258150.012	263533.896	269177.367	275061.275	281150.376	287395.046	293759.907	300236.786	306838.347	313577.399	320471.931	276146.014

Figure 8: Total Population Size of Each Language from 1950-2015

1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2017
77114679	80787402	84460125	93221637	101983149	127262761	152,542,373	160,700,028	172,604,257	190,531,600	220,019,266	247,585,744	257,715,425

Figure 9: International Migration Stock data from 1960-2015 [19]

	1970			1996			2010			2020		
	gdp/capita	gdp growth	gdp	gdp/capita	gdp growth	gdp	gdp/capita	gdp growth	gdp	gdp/capita	gdp growth	gdp
World	802.71	4.66	2.9578E+12	450.43	4.33	3.15527E+13	9516.35	4.31	3.13388E+13	12248.24	3.70	93496.302
English	918.83	6.31	1.36008E+12	918.83	3.61	6.90895E+11	17954.40	1.85	2.10308E+13	25027.41	4.59	2.96386E+13
Mandarin	258.70	16.59	94522482123	258.70	-9.57	60420929928	25565.10	12.94	6.33704E+12	58229.04	2.60	340.414
Hindi	183.41	8.93	61809679002	183.41	3.62	36648253452	2498.87	6.61	1.65976E+12	4614.42	5.60	3230.981
Arabic	299.28	9.43	43225319258	299.28	3.31	11614650518	12758.82	5.14	2.32756E+12	15752.64	3.42	3159.775
Russian	509.42	3.23	17086956522	509.42	1.16	13995067818	7465.43	5.52	2.50686E+12	8393.77	2.96	2938.906
German	1559.63	5.43	43821407448	1559.63	4.78	28478088857	82394.87	3.14	1.51752E+12	77158.65	1.95	1871.092
Japanese	479.00	-1.02	2.11514E+11	479.00	12.04	44307342950	44507.68	4.19	5.7001E+12	42049.62	0.20	5279.771
Punjabi	81.92	8.26	71616889369	81.92	4.86	40242980931	1192.96	5.93	1.83402E+12	2354.68	6.80	3224.87
Portugese	285.26	10.69	50581697897	285.26	7.91	18358770317	11301.77	5.88	2.58783E+12	7239.20	2.84	2820.729
Hausa	92.96	25.01	12545849083	92.96	0.19	4196092258	2327.32	7.84	3.69062E+11	2876.83	1.70	588.841
Spanish	393.15	5.41	1.69554E+11	393.15	4.70	52192024617	9049.93	3.94	4.48867E+12	11716.59	2.86	5351.954
Bengali	84.99	5.39	70582522329	84.99	4.89	40810818943	1051.72	7.92	1.7719E+12	2146.94	7.45	3551.26
Malay	248.03	8.34	67552560222	248.03	6.49	39156629329	23063.73	8.77	2.16176E+12	26009.69	6.08	3994.092
French	547.37	5.80	2.4726E+11	547.37	3.28	1.18756E+11	18253.55	3.65	5.0908E+12	16300.89	4.18	5930.16

Figure 10: data for compute language utilities

Appendix B: Computational Code

```
function z = home2_de2(Q, Q_fixed, u)
%Set all variables passed in from main program
k = Q(1);
K = Q(2);
t_data = Q_fixed(:,1); % time period data
C_data = Q_fixed(:,2); % total speakers data
t_f = t_data(end); % time length
P = 2525; % initial world population
C_0 = C_data(1);
L_2_0 = 371;
%Solve DE using current parameter values
[t,C]=ode45('population1_de',[1950:1:2015],[P C_0 L_2_0]',[],Q, u);
%Interpolate model to t-values from data set
C_model = interp1(t,C(:,2),t_data);
%Calculate 2-norm of error (i.e. RMSE, the square root of MSE)
z = norm(C_model-C_data);

function dy=population1_de(t,Y,flag,Q, u)
k= Q(1); % growth rate
K= Q(2); %environment capacity (Max population)
P= Y(1);
L= Y(2);
L_1= Y(3);

dy(1,1)= 0.025444583098593*P.*(1-P/(1.353911589882259e+04)); %population growth rate
dy(2,1)=  k*L.*(1-L/K) + dy(1,1)*(0.00001*P-0.012232)*u; %total language speakers growth rate
dy(3,1)=  k*L_1.*(1-L/K); % total native language speakers growth rate
```

Figure 11: optimize and ODE function

```
%% Load Data
load totalLanguage.mat;
load languageUtilities.mat; % mean3 variable as language utilites
load mean1.mat; % mean1 variable as language utilites from 1950 - 2015
load mean2.mat; % mean2 variable as language utilites from 2015 - 2070
data = table2array(totalpopulationS4);
num = size(data,1); % number of languages
data = data'/1000;
t_data = (1950:1:2015); % time interval
m = length(t_data); % length of time interval
%
native_data = [371 897 329 290 153 76 128 148 218 85 436 242 90 76];

%Initial parameter guesses
k = 0.01;
K = 3000;
Q = [k K];

% Using to contain Q_opt of each language
D = [];
% Using to contain the percentage of each language over population
percentage1 = []; %2050
percentage2 = []; %2015
native1 = []; %2050
native2 = []; %2015
world1 = [];
world2 = [];
% data for world population
P_data = [2525 2758 3018 3322 3682 4061 4440 4853 5320 5735 6127 6520 6930 7349];
P = 2525; % initial world population (1950)
```

```
for i = 1:num
    u_1 = mean1(i); % language utility 1950 - 2015
    u_2 = mean2(i); % language utility 2015 - 2070
    C_data = data(:,i);
    C_0 = C_data(1); % initial population (1950)
    N_0 = native_data(i); % initial # native speaker

    Q_fixed = [t_data' C_data];     %Variables that fminsearch must not change


    %Perform optimization (remember...no fractional order exponents)
    options = optimset('MaxFunEvals',10000);
    [Q_opt,z_opt]=fminsearch('home2_de2', Q, options, Q_fixed, u_1);
    D = [D; Q_opt];

    [t_1,C_1]=ode45('population1_de',[1950:1:2015],[P C_0 N_0]',[],Q_opt, u_1);
%     percentage1 = [percentage1; C(end,2)/C(end,1)];
%     percentage2 = [percentage2; C(66,2)/C(66,1)];
    [t,C]=ode45('population1_de',[2015:1:2070],[C_1(end,1) C_1(end,2) N_0]',[],Q_opt, u_2);
    world1 = [world1; C(end,1)];
    world2 = [world2; C(1,1)];
    percentage1 = [percentage1; C(end,2)/C(end,1)];
    percentage2 = [percentage2; C(1,2)/C(1,1)];
    native1  = [native1; C(end,3)];
    native2  = [native2; C(1,3)];
end;

percentage = [percentage1 percentage2];
native = [native1 native2];
```

Figure 12: Main model code

```
clear all;
close all;
%Load data
load L_utility.mat;
LanguageS1 = table2array(LanguageS1);
language_utility = LanguageS1(2:end,:);
world = LanguageS1(1,:);
m = size(language_utility,2);
%scale down to range 100
for i = 1:m
    language_utility(:,i) = rescale(language_utility(:,i),1,100);
end
X = [5; 1; 4]; % weight of GDP/capita, growth-rate, GDP respectively
% Calculate the language utility index of 1970-1996, 2012-2020
a_1970 = language_utility(:,13:15)*X;
a_1996 = language_utility(:,1:3)*X;
a_2012 = language_utility(:,10:12)*X;
a_2020 = language_utility(:,16:18)*X;
% Calulate the language utility (percentage)
percent = [a_1970 a_1996 a_2011 a_2020];
total = percent./sum(percent)*100;
mean1 = sum(total(:,1:2),2)/2;
mean2 = sum(total(:,3:4),2)/2;
```

Figure 13: language utilities compute

```
data = xlsread('migratiinnet.xlsx');
data = data';
t_data = data(:,1);
C_data = data(:,2)/1000000;
plot(t_data,C_data);

P_data = [3018 3322 3682 4061 4440 4853 5320 5735 6127 6520 6930 7349 7599];
P_1 = P_data;
P_2 = P_data.^2;
m = length(P_data);
P = [ones(m, 1) P_1' P_2'];

% Calculate the parameters from the normal equation
theta = normalEqn(P, C_data);

% Display normal equation's result
fprintf('Theta computed from the normal equations: \n');
fprintf(' %f \n', theta);
fprintf('\n');
```

```
function [theta] = normalEqn(X, y)
%NORMALEQN Computes the closed-form solution to linear regression
%   NORMALEQN(X,y) computes the closed-form solution to linear
%   regression using the normal equations.
theta = zeros(size(X, 2), 1);
theta = (X' * X)^(-1) * X' * y;

end
```

Figure 14: International migration stock compute