For office use only	Team Control Number	For office use only
T1 _____	**81393**	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	**B**	F4 _____

2018 Mathematical Contest in Modeling (MCM) Summary Sheet

Summary

Language is a set of sounds, written symbols and the system of communication in speech and writing used by people of a particular country of area. The number of human languages in the world is estimated to be more than 6900. This article is dedicated to analyzing and forecasting the distribution of multiple languages and analyzing the issue of decision making in internationalized locations.

For Part 1, we establish Exponential Growth Model and Gray Forecast Model GM (1,1). First, by collecting and organizing the data obtained, we consider the impact of national policy, economic, social pressure and the language(s) used in school on the model. We predict the number of native speakers in 16 languages and the number of speakers in the second language by using Exponential Growth Model, we draw the following conclusion. In the next few years, Bengali grew faster as native speakers while the number of Japanese and German as native speakers showed a decreasing trend. The number of people who use English as a second language still dominate while other languages will have different growth rates but not obviously. Then, we forecast the changes of languages in the next 50 years. The result of Exponential Growth Model prediction is ideal, but the result of Gray Forecast Model prediction is not satisfactory. So, we abandon the Gray Forecast Model. From the result of Exponential Growth Model, the top 3 usage is Chinese, English and Spanish. Finally, we consider the global population and human migration patterns effect on our model. By calculating the migration of countries over the years and using that to predict the geographical distribution of languages in the next 50 years. The result demonstrates that.

For Part 2, we use the ideal solution. Based on the result of Part I analysis and forecasting, the top 10 cities in terms of usage and the cities with the best economic growth among the 10 native speakers were selected as candidates and evaluated. Take the top 6 countries for the best assessment and use the 6 most economically desirable cities in the country as the recommended 6 new locations of office. In our result, the worker in these 6 offices at least have a good command of English and its native language while Spanish and native language is needed in long run. We recommend that the company open only 5 new locations based on the changes in the distribution of languages in the next 50 years.

Finally, we performed a sensitive analysis that considers the impact that different countries may take in the future on our model. And wrote a two-page memo to the company's COO to explaining our results and giving our advice.

Key words: Exponential Growth Model, Gray Forecast Model, Ideal Solution

The Distribution of Languages in Future

Contents

1 Introduction

1.1 Background

With the burgeoning development of economic growth, people in growing number can use two or more different languages to communicate with others. Hence, the total number of speakers of a particular language will be fluctuated by factors such as government in a country, the language(s) used in schools, social pressures, migration and assimilation of cultural groups, and immigration and emigration with countries that speak other languages, etc.

Along with the process of globalization and the development of worldwide multi-cultures, to explore the different languages has become a common value orientation for all countries. In order to have a better understanding future of the world, we need to predict the language changes with time and region.

1.2 Restatement of the Problem

To some degree, the development of time will change the language distribution. A complicated social and cultural phenomenon needed to know by predicting the changes in language which may proceed as follows:

- Consider the influencing factors and speculate on the distribution of users in various languages.
- Predict what will happen in the next 50 years on the numbers of native speakers and total language speakers.
- Predict what will happen in the next 50 years while the global population and human migration patterns.

A company would like to be international by opening 6 additional international offices but there are several issues needed to be resolved.

- Where to locate these office and which language to use in it.
- What will happen if open less than 6 offices.

1.3 Literature Review

According to the data[10], we can know that there are Exponential Growth Model of the current forecasting methods for population and the Gray Forecast Model. From the literature, we can see that the gray forecasting model is not suitable for long-term forecasting. Therefore, in our dissertation, the models compare the long-term prediction results to select a more suitable long-term prediction model.

By adoption[9], we have probably learned about the worldwide implications of some of the languages as a second language and are considered in our dissertation.

As a result, we have a general understanding of some of the policies[5][6][7][8] that Germany and India have taken on the size of the population, and the post-implementation impact that we took into account in our sensitivity analysis.

1.4 Data Collection and Analysis

We have found the number of users of each language over the years as a native language and a second language through the[13], but we only consider 16 of them in the dissertation. So, we sort out the 16 languages we need from the original data and use this in the dissertation to predict. But the data in *Ethnologue* [13]only provides the number of languages used as the second language in 2010 and 2015, the gray model can only predict the distribution at 5-year intervals.

Through the CIA[1] we can find the population over the years and the net migration rate per 1000 population. We consider these data as an immigration factor in the paper and discuss the impact on the original results.

2 Terminology and Definitions

Parameter	Definition
$x^{(0)}$	Reference data column of GM (1,1)
$x^{(1)}$	Accumulate the sequence of GM (1,1)
$z^{(1)}$	The average generates a sequence of GM (1,1)
$\boldsymbol{\varepsilon}(k)$	Residual
$\boldsymbol{C}^*$	The positive ideal solution
$\boldsymbol{C}^0$	The negative ideal solution
c_j^*	The *jth* property value of positive ideal solution $\boldsymbol{C}^*$
c_j^0	The negative *jth* property value of the negative ideal solution $\boldsymbol{C}^0$
s_i^*	The distance from the candidate city to the positive ideal solution
s_i^0	The distance from the candidate city to the negative ideal solution
f_i^*	the rank index of each city

3 Problem Analysis

Due to the workload of collecting all the data in 26 languages listed on the subject, we only predicted the distribution of 16 languages.

For Part I, we use the exponential growth model and the GM (1,1) model to make a short-term forecast of the distribution in 16 languages, and use the exponential growth model for long-term prediction.

For Part II, use the best-performing cities in the top ten languages of the country for the candidate office locations, and use the ideal solution to find the overall rating index for each city, taking the top 6 as the new office locations. And analyze the difference between short-term and long-term.

For Part III, we wrote a two-page memo to coo, given and explained our suggestions.

4 Model Assumptions

- Ignore the occurrence of catastrophic events (asteroid collisions, destruction of the earth);
- Ignoring the influence of immigrants' descendants on native language usage;
- Neglecting the impact of countries with a low population on the use of languages;
- Neglecting the impact of offspring born in different countries on native language usage by couples;
- Ignoring the changes in population that may result from the greenhouse effect;
- No massive war will take place in the next 50 years resulting in a sharp drop in the population.

5 Models and Solutions

5.1 For Part I

The List of Language by Total Number of Speakers in Wikipedia list 26 languages which numbers of users ranged from 66 million to 1.09 billion. Due to the complexity, time-consuming and complicated workload of surveying the number of users and the geographical distribution of the 26 languages over the years, we only consider using languages rank in the top 16[13]. These languages are list as follow:

Table 1 Top 16 Language

Rank	1	2	3	4	5	6	7	8	9
Language	Chinese	Spanish	English	Hindi	Portuguese	Bengali	Russian	Japanese	Javanese
Native	897	337.3	338.2	262	181	201.5	141.5	121.5	85
Rank	8	9	10	11	12	13	14	15	16
Language	Japanese	Javanese	German	Korean	French	Telugu	Marathi	Turkish	Tamil
Native	121.5	85	86.9	72.5	70.1	70.1	68.2	71	69

We use Exponential Growth Model and Gray Forecast Model to predict the number of 16 languages user's distribution over time. The number of users includes the number of native speakers and second language.

To begin with, we don't consider any other factors in order to establish 2 models to predict the number distribution of native language and second language in the next few years.

5.1.1 Exponential Growth Model

We defined a exponential growth function:

$$y = ae^{bx}\text{-}c$$

where a, c is dynamic parameter, b is the rate of increase or decrease

Assuming that all people in a country speak their native language, so that the number of native speakers will change with the number of the country's population. Through the data of native language in recent years, using the cftool toolbox in MATLAB to fit these 16 languages of its exponential growth function and calculate the value of a, b, c.

Table 2 Value of a,b,c

Value \ Language	Chinese	Spanish	English	Hindi	Portuguese	Bengali	Russian	Japanese
a	490.3	6.666	1192	-924	2.647	3426	-4334	-3.293
b	0.002645	0.0694	0.0015	-0	0.05637	0.0003	0.0001	0.01678
c	-382.3	-314.2	886.1	-1183	-173.4	3241	-4485	-126
Value \ Language	Javanese	German	Korean	French	Telugu	Marathi	Turkish	Tamil
a	2745	-17.19	3.621	8.213	0.03551	0.1837	24.27	4.21
b	0.000246	0.025	0.0569	0.031	0.1426	0.0465	0.0357	0.03377
c	2671	-113.4	-62.8	-55.9	-69.64	-67.79	-25.34	-61.43

Part of the fitting figure is as follows:

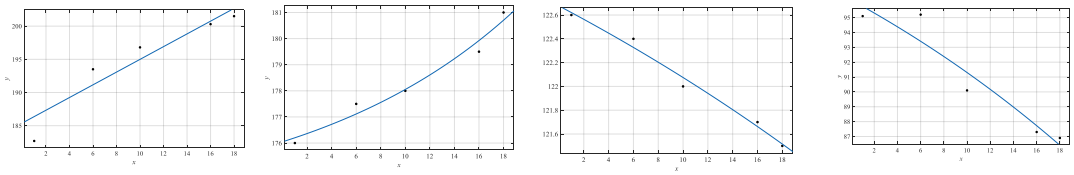

Figure 1 Curve fitting

This can predict the number of native speakers in these 16 languages over the next 8 years which shows in the following table.

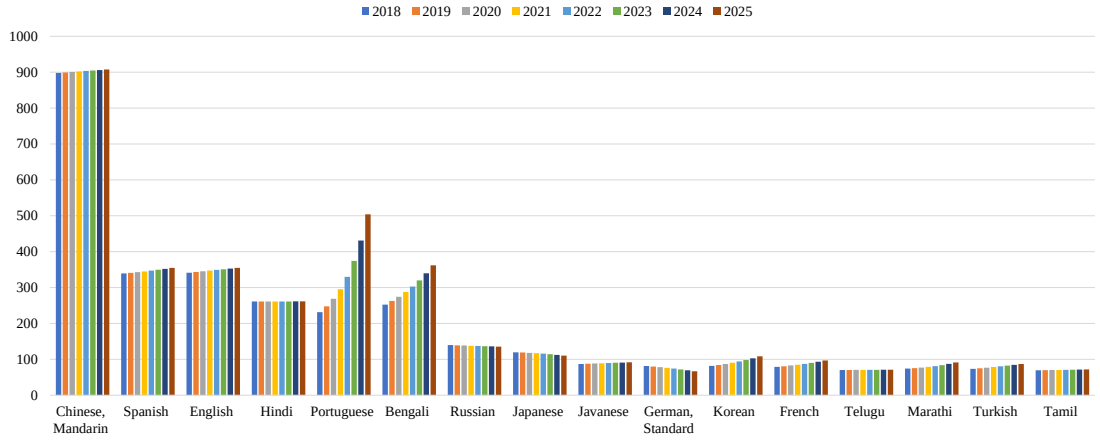

Figure 2 Growth trend of 16 native language

The table represents that Portuguese and Bengali will have a faster growth trend in their usage in the next few years as their mother tongue. The rapid population growth in Portugal and India in recent years has directly led to a rapid increase in the number of native speakers in both languages. Besides, Japanese and German showed a declining trend due to the negative growth trend of the population in Japan and Germany in recent years and the corresponding reduction in the number of people who use as a native language.

Using the same predict method above, we use the data of these 16 languages as their second language in 2015 to 2017 to predict the changes in next 5 years.[13]

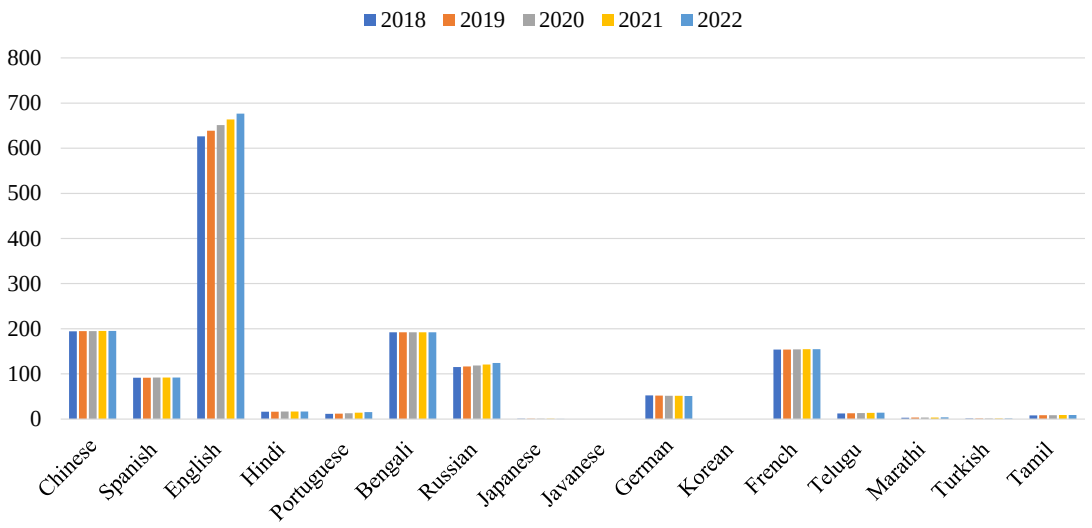

Figure 3 Growth trend of 16 second language

Because we can't find Korean and Javanese these two languages as the number of second languages, it is not displayed. The table indicates that English is still the most commonly used language in the world as a second language but with little change in other language.

5.1.2 Gray Forecast Model

Gray Forecast Model[12] is one of the common predicting model which is through the accumulation of the original data generated by the approximate exponential law and then predict the method. We use the predict model GM (1,1) to predict the data we got.

Put these 16 languages as native language can establish a reference series

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2),\cdots, x^{(0)}(n))$$

And the cumulative generation sequence (1-AGO) for one time is

$$\begin{aligned} x^{(1)} &= (x^{(1)}(1), x^{(1)}(2),\cdots, x^{(1)}(n)) \\ &= (x^{(0)}(1), x^{(0)}(1) + x^{(0)}(2),\cdots, x^{(0)}(1) +\cdots+ x^{(0)}(n)) \end{aligned}$$

where

$$x^{(1)}(\mathrm{k}) = \sum_{i=1}^{k} x^{(0)}(i),\ k=1, 2,\cdots, n$$

The mean generation sequenced of $x^{(1)}$ is

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3),\cdots, z^{(1)}(\mathrm{n}))$$

where

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k) + 0.5x^{(1)}(k\text{-}1),\ k=2, 3,\cdots, n.$$

Use it to establish the gray differential equation can get

$$x^{(0)}(k) + az^{(1)}(k) = b,\ k = 2,3,\cdots, n$$

Where a is the development coefficient and the size of a determines the development trend of $x^{(1)}(k)$ and $x^{(0)}(k)$. If the result is negative, then the sequence shows a trend of increasing and the greater the absolute value of a, the faster growth. If the result is positive, then the sequence shows a declining trend and the greater the absolute value of a, the faster the decay.

b is gray role. As a system, its role should be exogenous and can't be directly observed. Its size represents the change of sequence, its change represents the change of characteristic behavior. b in the system is equivalent to the role of the amount of its exact content and meaning

of gray so called gray role.

The corresponding albino differential equation is

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b$$

Suppose

$$\boldsymbol{u} = [a,b]^{\mathrm{T}},$$

$$\boldsymbol{Y} = [x^{(0)}(2), x^{(0)}(3),\cdots, x^{(0)}(\mathrm{n})]^{\mathrm{T}},$$

$$\boldsymbol{B} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix},$$

the estimated value of the u that makes

$$\boldsymbol{J}(\boldsymbol{u}) = (\boldsymbol{Y}\text{-}\boldsymbol{B}\boldsymbol{u})^{\mathrm{T}}(\boldsymbol{Y}\text{-}\boldsymbol{B}\boldsymbol{u})$$

and the minimum value is

$$\hat{\boldsymbol{u}} = [\hat{a}, \hat{b}]^{\mathrm{T}} = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}$$

So, we can get the albino differential equation is

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{\hat{b}}{\hat{a}})e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}, \ k = 0, 1,\cdots, n\text{-}1,\cdots$$

In order to ensure the feasibility of the gray prediction method, we need to test the obtained result.

Suppose the residual is

$$\boldsymbol{\varepsilon}(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}, \ k = 1,2,\cdots, n$$

where

$$\hat{x}^{(0)}(1) = x^{(0)}(1)$$

if $\boldsymbol{\varepsilon}(k) < 0.2$, it can regard that meet the general requirements.

If $\boldsymbol{\varepsilon}(k) < 0.1$, it can regard that reach the higher requirements.

By means of the above method, the solution to the albino differential equation for the usage of the native languages in 16 languages is obtained:

Table 3 Equation (Shown at Appendices 4)

Language	Equation
Chinese	$119600e^{0.007332t} - 118700$
Spanish	$17222e^{0.01855t} - 16900$
…	…
Turkish	$448e^{0.1145t} - 397.5$
Tamil	$3159e^{0.02071t} - 3093.0$

From

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k)$$

we can get the prediction of the usage of these 16 languages as native language in 2020 and 2025.

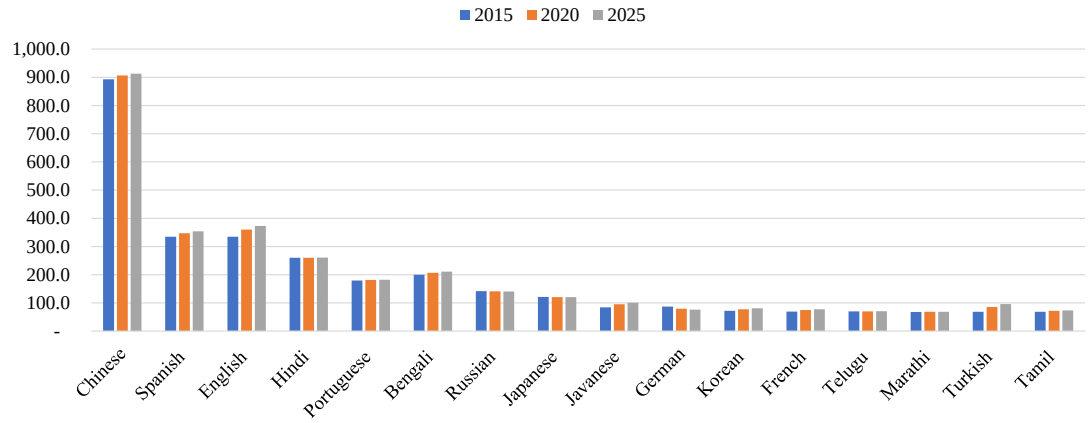

Figure 4 Growth trend of 16 native language

The table indicates that the growth trend of Portuguese and Bengali is obvious and the trend of Japanese and German is declining.

Using the same method to predict these 16 languages as their second language distribution in the next 5 years. The result is shown as follows.

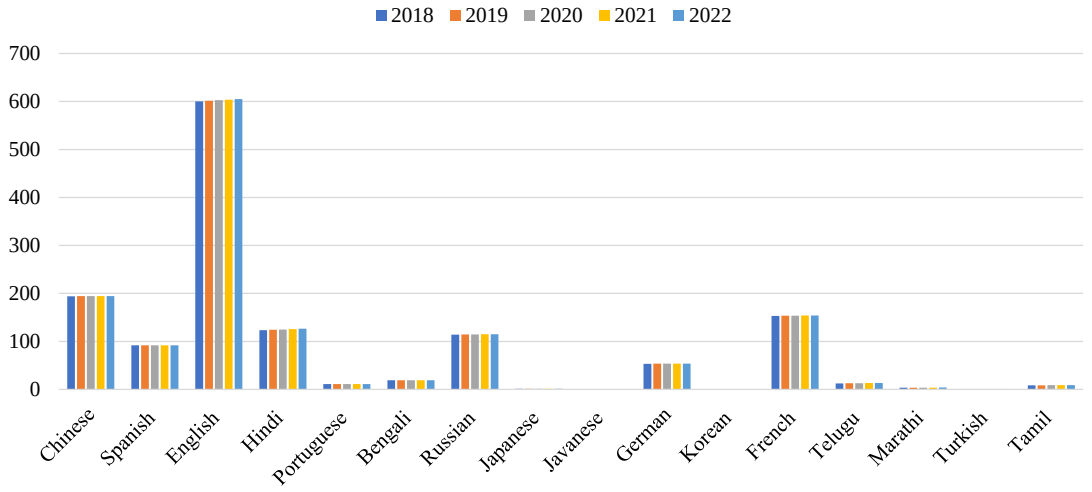

Figure 5 Growth trend of 16 second language

It can also be seen that the use of English as a second language still dominates over the next 5 years while other languages is with little change in the growth. This is in line with the conclusion reached earlier using exponential growth.

To sum up, Portuguese and Bengali grew faster as native speakers in the coming years while the number of Japanese and German as native speakers showed a decreasing trend. In the second language, English will still dominate in the coming period while other languages will have different growth rates but not obviously.

Though we predict the usage of 16 languages as native language and the usage change in the next few years while as the second language but without considering any other influencing factors. Obviously, this predict is not real and reality.

Based on the original model, we start to consider the national policy, economic, social pressure, the language(s) used in schools and other factors and analyze the impact of these factors on the existing model.

5.1.3 Optimization of Exponential Growth Model

In our exponential growth model, *b* represents for the rate of increase and decrease. So,

the influencing factors above on the original model is the impact on the rate *b*.

We have assumed that the use of a language as a native language is the same as the national population growth trend. From the above influencing factors, national policy and economic have a great impact on the number of the country. After searching the research can show that China is completely liberalized its second child policy in 2015. Besides, China's GDP has been growing rapidly in recent years. All of these factors will cause the rate of population growth in China to change. However, we can't accurately know the specific impact of these changes on the rate. We introduced the coefficient α to roughly assess the effect of these factors on the growth rate *b* and quantified in exponential growth function.

$$y = ae^{abx}\text{-}c \tag{1}$$

We pass the influence factors in the weight assignment between 0 to 1, and then add the value of α. With more information [1][2][3][4], we derive the weight values in 16 languages and the corresponding α values:

Table 4 Weight of 16 language (Shown at Appendices 5)

Language / Weight	Chinese	Spanish	…	Turkish	Tamil
Policy	0.75	0.45	…	0.45	0.6
Economic	0.8	0.35	…	0.4	0.5
Total	1.55	0.8	…	0.85	1.1

Put the value above into formula 1 we can predict the number of distribution in the next 8 years.

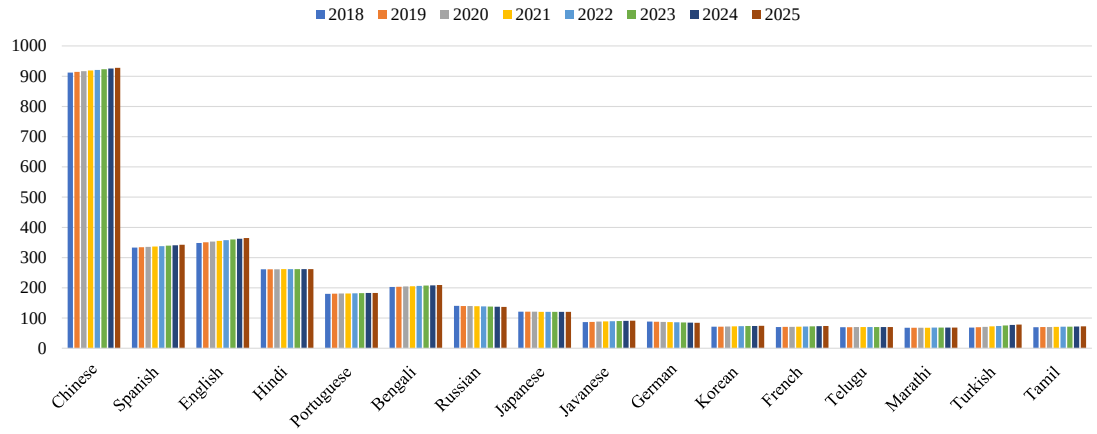

Figure 6 Growth trend of 16 native language

The figure shows that under the consider of national policy and economic, Portuguese and Bengali slowed down. The growth rate of Chinese, Spanish and English showed a rapid growth. The decline rate of Japanese slowed down while that of German remained relatively fast downtrend. Other languages have not changed much.

When is come to put these 16 languages as the second language and use the same method, we consider the distribution of 16 languages as a second language by 4 factors of national policy, economic, social pressure, and the language(s) used in school. Similarly, the weight assignments for the influencing factors and the corresponding α values are as follows:

Table 5 Weight of 16 language (Shown at Appendices 6)

Language / Weight	Chinese	Spanish	…	Turkish	Tamil
Policy	-0.2	0.2	…	0.15	-0.1
Economic	0.8	0.4	…	0.35	0.4
Social	0.4	0.2	…	0.3	0.3
School	0.5	0.3	…	0.3	0.2
Total	1.5	1.1	…	1.1	0.8

Using Equation 1, we can predict the number distribution over the next 8 years.

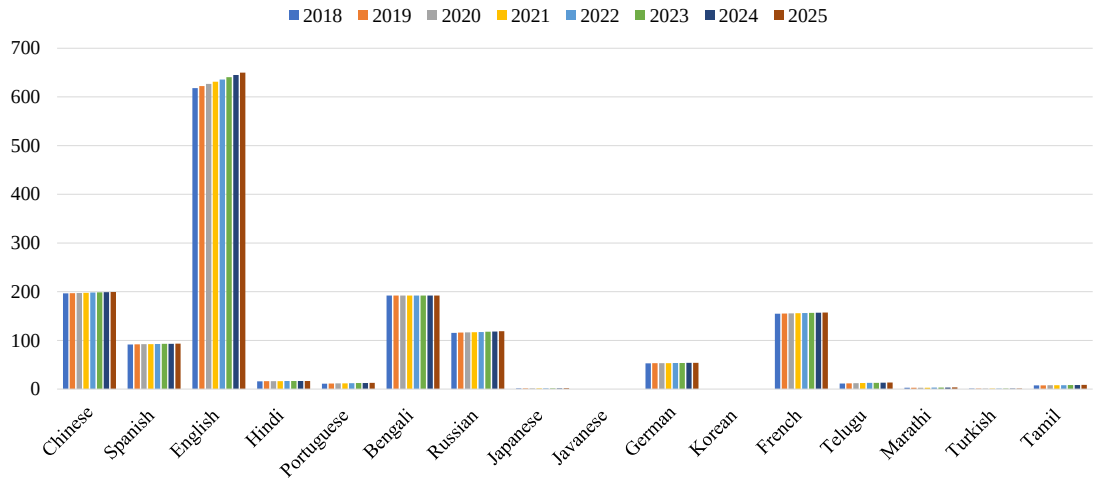

Figure 7 Growth trend of 16 second language

It can be seen that the growth trend of English as a second language is still obvious when the national policy, economy, social pressure and the language(s) used in schools are taken into consideration, while the changes in other countries are not significant.

To sum up, we can come to 16 languages in the next 8 years the number of distribution:

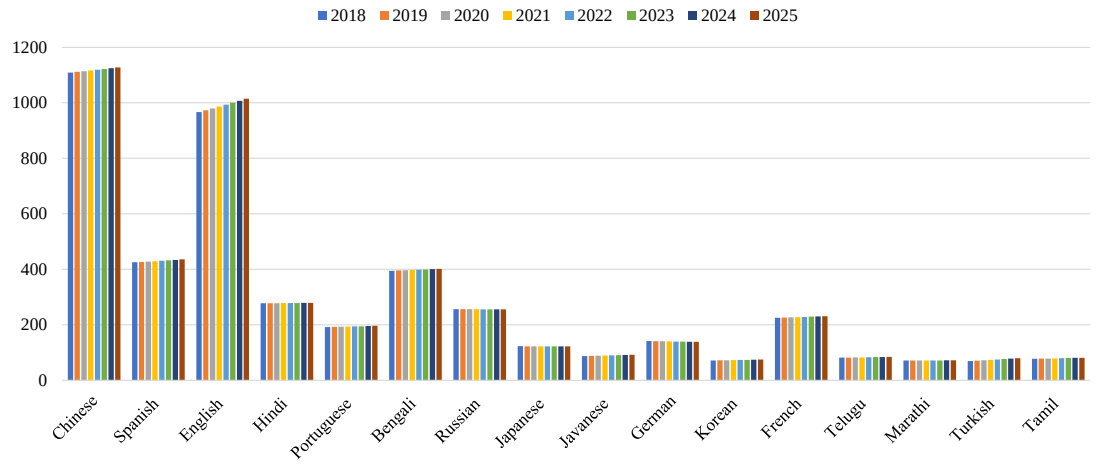

Figure 8 Growth trend of total language

We can conclude that in the next eight years, the number of people using Chinese is the largest, followed by English which followed by Spanish, and little change in other languages.

5.1.4 Optimization of Gray Forecast Model

In the GM (1,1) Model, a is the development coefficient which is the rate of growth. So, the influencing factors of GM (1,1) is based on the change of the development coefficient a. Similarly, we can't accurately know the specific changes of these factors to the coefficient of development. We introduce β coefficients to roughly assess the effect of these factors on the coefficient of development β and quantified in GM (1,1) Model.

$$x^{(0)}(k) + a\beta z^{(1)}(k) = b,\ k = 2,3,\cdots,\ n \tag{2}$$

The influence factors are weighted between 0 and 1 and the addition is the value of β. Through the research, we can get the weight value of the 16 languages and the corresponding value β.

Table 6 Weight of 16 language (Shown at Appendices 7)

Language / Weight	Chinese	Spanish	…	Turkish	Tamil
Policy	0.75	0.45	…	0.45	0.6
Economic	0.8	0.35	…	0.4	0.5
Total	1.55	0.8	…	0.85	1.1

Using formula 2, we can predict the number distribution of 2020 and 2025

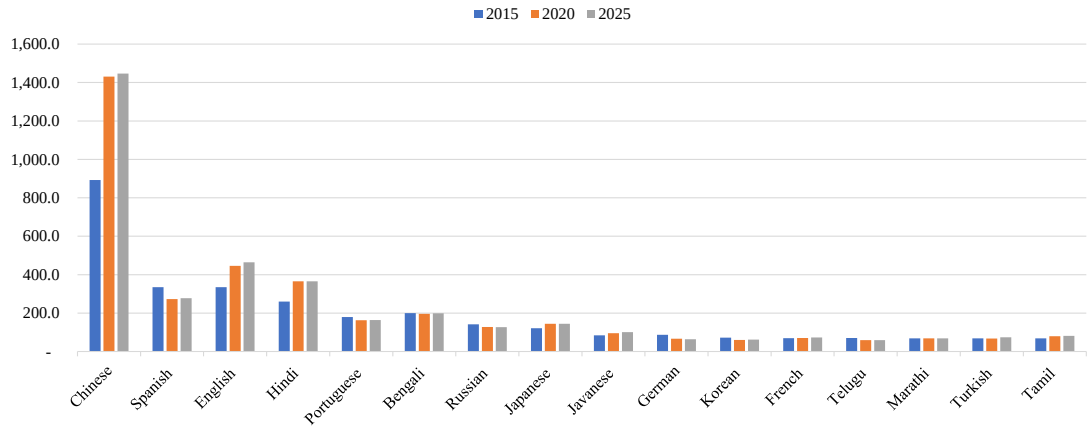

Figure 9 Growth trend of 16 native language

It can be seen that Chinese is the most frequently used as a native language in the future, followed by English and Hindi, roughly in line with the results predicted by the Exponential Growth Model.

Taking the national policy, economic, social pressure and the language(s) used in school into account, we can optimize the predictions of 16 languages as the second language. The weight assignment of each factor and the corresponding β values are as follows:

Table 7 Weight of 16 language (Shown at Appendices 8)

Language / Weight	Chinese	Spanish	…	Turkish	Tamil
Policy	-0.2	0.2	…	0.15	-0.1
Economic	0.8	0.4	...	0.35	0.4
Social	0.4	0.2	…	0.3	0.3
School	0.5	0.3	…	0.3	0.2
Total	1.5	1.1	…	1.1	0.8

Using formula 2, we can predict the number distribution of 2020 and 2025.

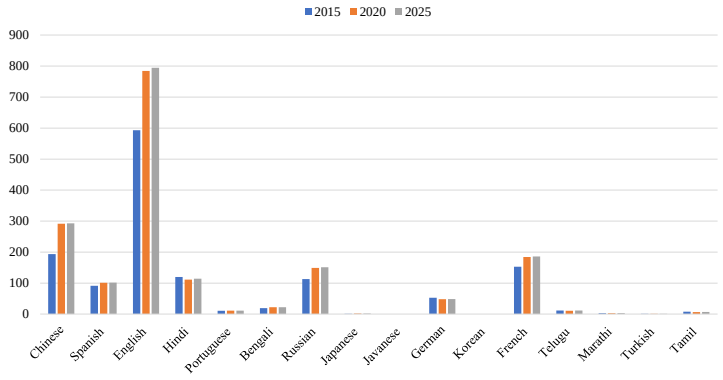

Figure 10 Growth trend of 16 second language

The figure shows that English as the second language is being dominated, followed by Chinese and French, the result is roughly the same with the Exponential Growth Model.

To sum up, we can get the number distribution of 16 languages in 2020 and 2025.

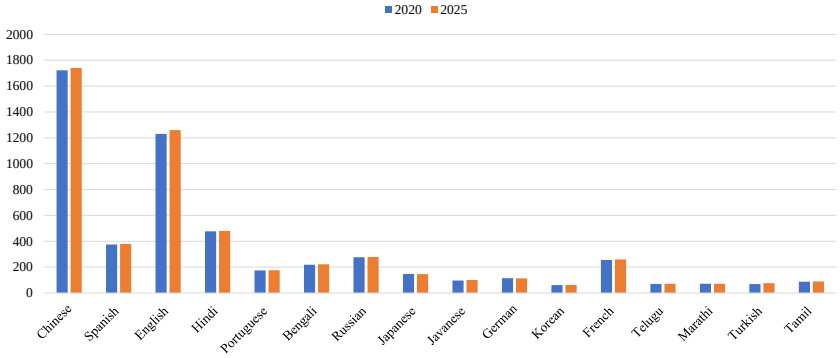

Figure 11 Growth trend of total language

Compare the result, we can see that under the influence of four factors of policy, economy, social pressure and school, Exponential Growth Model and Gray Forecast Model GM (1,1) can predict the number distribution of 16 languages in the next few years while the situation is roughly the same. The result is: the largest number of people use Chinese, followed by English, little change in other languages

For Problem B of Part I, we use the optimized exponential growth model and the Gray Forecast Model GM (1,1) to predict the number of people who use these 16 languages as their native language respectively in the next 50 years (including 2018) and the total number of users. The result is as follows:

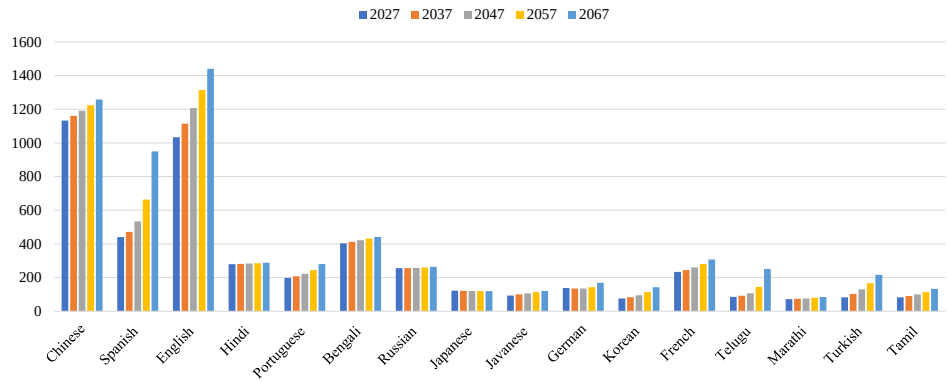

Figure 12 Growth trend of 16 total language

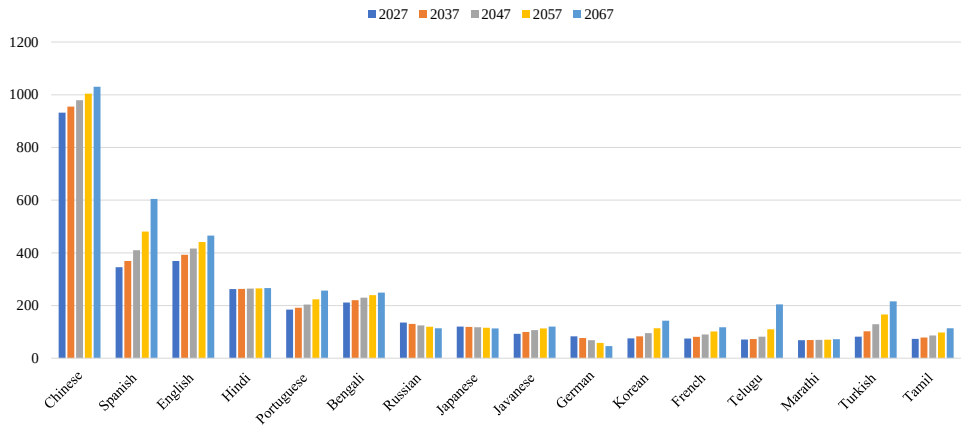

Figure 13 Growth trend of 16 native language

From these two graphs, we can see that the Gray Forecast Model GM (1,1) is not ideal for the next 50 years, so we discard the GM (1,1) model and the result of the Exponential Growth Model predicts the next 50 years is ideal.

In the short term (before 2037), the top ten languages of native speakers are Chinese, English, Spanish, Hindi, Bengali, Portuguese, Russian, Japanese, Turkish, Javanese. The total number of top ten Chinese, English, Spanish, Bengali, Hindi, Russian, French, Portuguese, German, Japanese.

In the long run, the number of speakers of English as native speakers will fall from 2nd to 3rd in the next 50 years, while Spanish will rise from 3rd to 2nd. In terms of the total number of users, English becomes the most used language instead of Chinese. French overtakes Hindi as fifth, Portuguese overtakes Russian as seventh, Telegu overtakes German as ninth.

These changes have occurred because the number of people who use English is growing year by year, while the natural population growth rate of countries where English is the native language is on the rise. The government has adopted a policy of encouraging birth, and the economic strength of these European countries has led to More and more people learn English. However, the natural rate of population growth in Russia has been declining year by year, resulting in a year-on-year decline of those who use Russian. After considering these factors, the results of our model prediction are appeared.

For the Problem C in Part I, we increased our consideration of the global population and human migration patterns.

Through the research, we can know the migration rate and the number of people over the years in which the native languages of these 16 languages.[13]

Table 8 Rate of immigration and total population (Shown at Appendices 9)

Language	Country	Rate of Immigration (migrants/1000 population) and Total Population (million)							
		2012	Total	2013	Total	2015	Total	2016	Total
Chinese, Mandarin	China	-0.34	1330.14	-0.33	1343.24	-0.44	1349.59	-0.32	1355.69
Spanish	Spain	2.73	46.51	5.02	47.04	8.31	47.37	7.24	47.74
	Argentina	0	41.34	0	42.19	0	42.61	0	43.02
…	…	…	…	…	…	…	…	…	…
Turkish	Turkey	0.53	77.80	0.5	79.75	2.16	80.69	0.46	81.62
Tamil	India	-0.05	1173.11	-0.05	1205.07	-0.04	1220.80	-0.05	1236.34

If the net immigration rate is positive, indicating that more immigrants moved into the country than the number of immigrants out of the country. If the net immigration rate is negative, immigrants went out more than the number of immigrants.

Similarly, using the Exponential Growth Model to predict the distribution of these 16 languages as native language over the next 50 years.

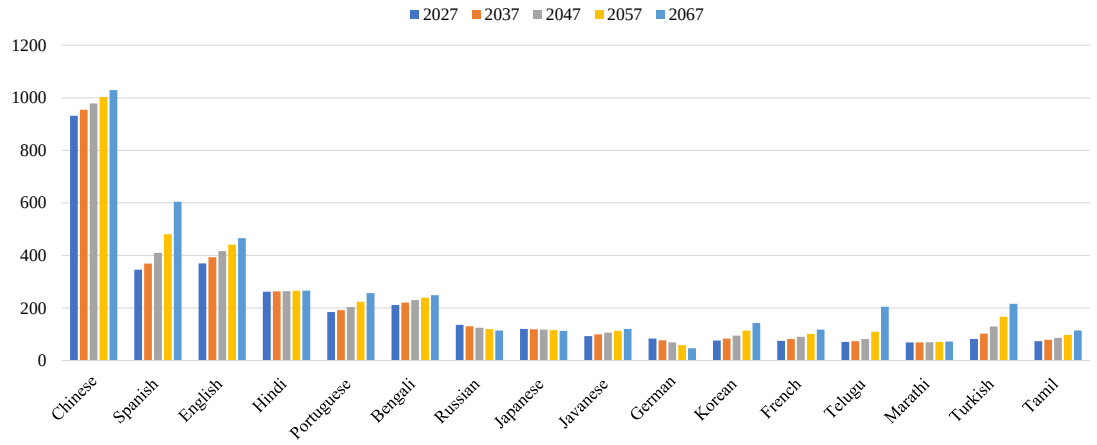

Figure 14 Growth trend of 16 native language

It can be seen from the figure that under the circumstances of considering immigrants, the use of Chinese as the mother tongue is still the largest. Spanish will surpass English in about 2047. This is roughly the same result as before.

The 20 countries are spread over 5 continents (Asia, Europe, North America, South America and Oceania) so we can know the geographical distribution of these 16 languages on 5 continents in the next 50 years as native languages:

Figure 15 The geographical distribution of language

According to result, we can predict the number distribution of these 16 languages as second language in the next 50 years.

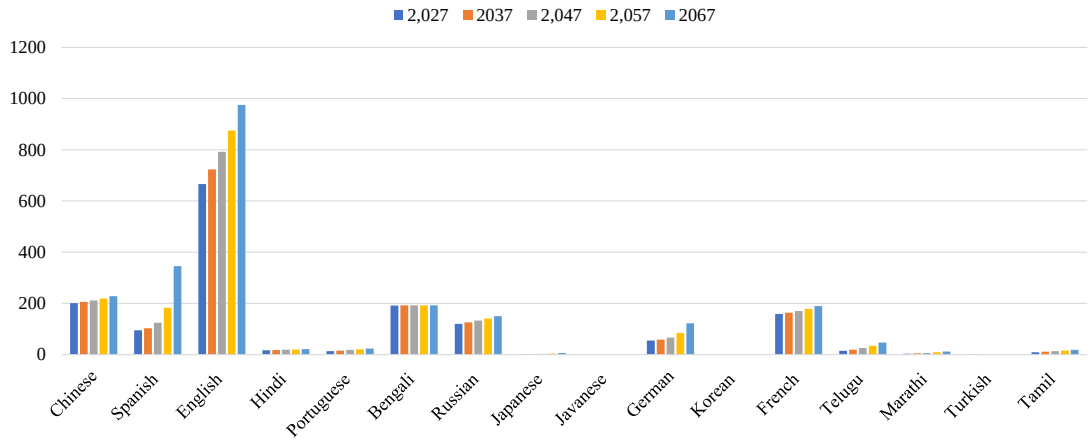

Figure 16 Growth trend of 16 native language

In summary, we can conclude that in the next 50 years, English will dominate as a second language, while Spanish will overtake Chinese as second, with little change in other languages.

5.2 For Part II

In the view of the company's need to select 6 cities around the world as a new office location, we choose the top 10 languages in total usage language and select the country that has the largest number of its native speakers. Use the ideal solution to evaluate and compare the 10candidate office locations, select the best evaluation of the first 6 as a new office location.

5.2.1 Ideal solution

We evaluate these 10 cities from 4 aspects: the economy (the annual amount allocated to the city by the country), the number of languages spoken, the level of traffic developed (several subway stations in the city), and the happiness index.

Specific steps are as follows:[10]

- Use the vector programming method to get the normalized decision matrix. Suppose decision matrix $\boldsymbol{A} = (a_{ij})_{m\times n}$ and the standardized decision matrix $\boldsymbol{B} = (b_{ij})_{m\times n}$.

  Where

$$b_{ij} = a_{ij} / \sqrt{\sum_{i=1}^{m} a_{ij}^2} \ , i = 1,2,\cdots, m; \mathrm{j} = 1,2,\cdots, n$$

- Construct a matrix of weighted norms $\boldsymbol{C} = (c_{ij})_{m\times n}$, suppose the weight vector as

$$\boldsymbol{w} = [w_1, w_2,\cdots,w_n]^{\mathrm{T}}$$

So

$$c_{ij} = w_j \cdot b_{ij},\ i = 1, 2,\cdots, m; j = 1, 2,\cdots, n$$

- Determine positive ideal solution $\boldsymbol{C}^*$ and negative ideal solution $\boldsymbol{C}^0$. Suppose the *jth* property value of positive ideal solution $\boldsymbol{C}^*$ is c_j^* and the negative *jth* property value of the negative ideal solution $\boldsymbol{C}^0$ is c_0^*.
- Calculate the distance between positive ideal solution and negative ideal solution in each city. The distance from the candidate city to the positive ideal solution is

$$s_i^* = \sqrt{\sum_{j=1}^{n} (c_{ij} - c_j^*)^2} \ , i = 1, 2,\cdots, m$$

- The distance from the candidate city to the negative ideal solution is

$$s_i^0 = \sqrt{\sum_{j=1}^{n} (c_{ij}-c_j^0)^2}\ ,\ i = 1, 2,\cdots, m$$

- Calculate the rank index of each city (Comprehensive evaluation index)

$$f_i^* = s_i^0/(s_i^0+s_i^*)\ ,\ i = 1, 2,\cdots, m$$

- According to f_i^* from descending to rank the pros and cons of each city.

5.2.2 For Problem A

For the Problem A, the top 10 languages predicted by Part 1 for the total number of users of the language in the short term (2037) are: London, Chinese, Spanish, Bengali, Hindi, Russian, Portuguese, Turkish.

Select the most economically developed cities among the 10 native speakers as candidate offices:

Table 9 City of 10 country

Country	English	Chinese	Spanish	Bengali	Hindi
City	London	Shanghai	Madrid	Dhaka	New Delhi
Country	French	Russian	Portuguese	German	Turkish
City	Paris	Moscow	Lisboa	Berlin	Ankara

Through the data, we create a candidate city evaluation data sheet:

Table 10 City evaluation data sheet (Shown at Appendices 10)

City	GDP	Language	Traffic	Happiness Index
London	$2.57	6	9	6.725
Shanghai	$12.24	7	8	5.245
…	…	…	…	…
Berlin	$3.65	6	6	6.994
Ankara	$0.84	7	6	5.389

Using MATLAB software to calculate the overall evaluation index of the candidate cities are as follows:

Shanghai, Berlin, London, Paris, New Delhi, Moscow, Madrid, Ankara, Dhaka, Lisboa.

Shanghai is in the top six cities, which is duplicates the company's current office location. So, we take the latter six cities as the new office location.

Similarly, the top 10 languages predicted by Part I for the total number of users of the language in the long term (2067) are: English, Chinese, Spanish, Bengali, French, Hindi, Portuguese, Russian, Turkish, German.

It is clearly show that, the long-term usage rankings are inconsistent with the short term, but no new countries appear on the list, so we can directly draw the nomination to London, Madrid, Dhaka, Paris, New Delhi and Lisboa.

In summary, we can see that in the short term, it is better to choose Berlin, London, Paris, New Delhi, Moscow and Madrid as the new office location, and in the long run, to choose London, Madrid, Dhaka, Paris, New Delhi and Lisboa as the new office space it is better. From the results, the 6 cities selected have a great influence on the world, and the economic strength

is very strong, so the result is more satisfactory.

5.2.3 For Problem B

For Problem B, we find from the results that Berlin only performed better overall in the short term but will be replaced by Dhaka in the long term, and Dhaka will have a better future trend than Berlin, so we suggest short-term establishment of new office locations can be less than 6, only 5.

However, these results are only based on the fuzzy forecast of our current model. The factors we consider are not comprehensive enough and the information we find is limited cause that it is inevitable that we will make erroneous judgments.

In order to select the most suitable new office location, based on the existing data, we also need to know the following items.

The exact destination and distribution of the number of immigrants from all over the world. With this data, we can consider in more detail the geographical distribution of languages as a second language.

The number of languages used over the year as a native language and second language. Since we have only found the data in recent years, there is a big error in our forecast results. If there are more years of data, we can ensure the accuracy of the forecast results.

The specific policies adopted by countries for population growth in the past years and their policies on indigenous languages and foreign languages. We have only found the policies of the past few years. These policies are only promulgated in recent years. If there are more relevant policies, then our weighting of policy considerations will be more reasonable and the forecast result will be more satisfactory.

6 Sensitivity Analysis of Languages

Since the data used in our prediction is not complete enough, the prediction results will inevitably show minor flaws. In this paper, we do not consider the possible future measures of population growth, as well as possible national and foreign languages, so the predicted results are compared single.

Therefore, we conducted a sensitivity analysis of the Exponential Growth Model.

We have reconsidered the policies or other measures that several countries may enact on future population and language changes and re-predicted.

6.1 Russia

The population of Russia has shown a negative growth trend in recent years. Although the government has implemented a series of policies to encourage childbearing, its success has not been much achieved. If the population of Russia still maintains the trend of negative growth and the trend is bound to be greater and greater, the population of Russia may be lower than 1 billion in the near future. Although the number of people who are learning Russian is on the rise, the number of people learning Russian will also decrease as a result of the decrease in the number of natives. Based on this, we think that Russia will step up efforts to encourage childbearing in the future.

Therefore, we adjusted the assignment of weights so that the coefficient α changed accordingly and re-predicted the future distribution of Russian:

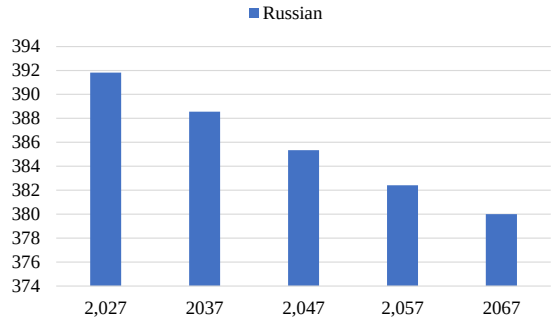

Figure 17 Growth trend of Russian

As can be seen from the table, the distribution of Russian will still show a downward trend, but has been flat than before.

6.2 Germany

Similar to Russia, the population of Germany has also shown a negative growth trend in recent years. Although Germany has introduced the incentive policy [5][6]: Women give birth to children receive substantial monthly subsidies. However, there is still not much help for reducing the population. Therefore, we can still assume that Germany will also increase its reward for childbearing in the future. In the same vein, the German language is predicted taking these factors into account:

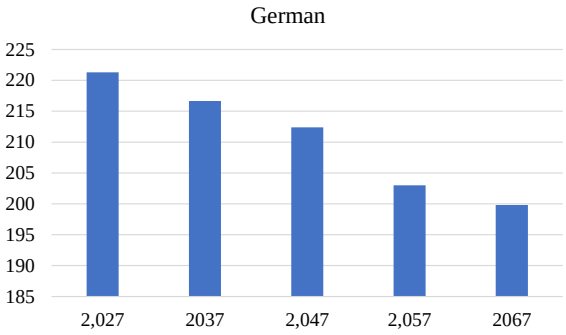

Figure 18 Growth trend of German

It can be seen from this that the number of German users has not dropped sharply before.

6.3 India

Through our projections, there is an increasing number of native speakers in India. This is because the increasing trend of population in India in recent years has led to an increase in the number of native speakers. Although India has long put forward the "birth control" policy[7][8], its policy on population growth has been ineffective due to its complicated political nature and its inability to implement it well. However, if India's population is allowed to maintain its current growth trend and India will become overcrowded, there will be insufficient consumption of environmental resources and various diseases.

Based on the above considerations, the same revaluation of several mother tongue in India, the forecast results are as follows:

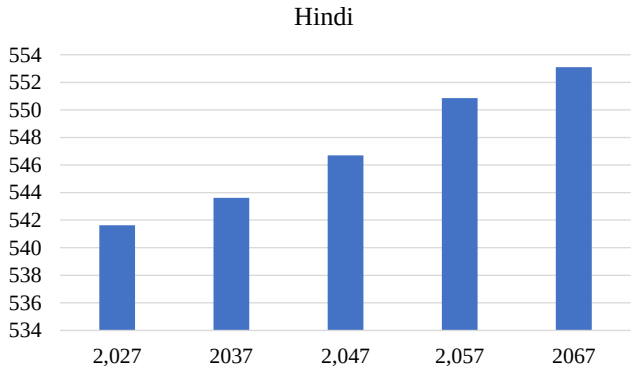

Figure 19 Growth trend of Hindi

The population of India is growing, but the rate of growth has slowed.

7 Strengths and Weaknesses

7.1 Strengths

- Using exponential growth model and Gray Forecast Model GM (1,1) two models to predict the distribution of each language in the short term, the result is quite satisfactory.
- Using the exponential growth model to predict the long-term distribution, the result is satisfactory.
- Predict the future distribution of languages to consider more factors, the results more in line with the actual situation.
- Reasonably use the ideal solution to select the more satisfied with the 6 new office locations.
- Through the analysis of short and long term, given the new office less than the client needs.

7.2 Weaknesses

- We can't know the complete data we need over the years through the data, so there is a small flaw in the forecast result
- For time reasons, the distribution of the 26 languages listed in the title is not fully predicted and only 16 languages are foreseen
- Because most of the data are quoted from the Internet, we can't guarantee the accuracy of the data, so the forecast result may have small deviations.

8 Future Work

Using the exponential growth model and the Gray Forecast Model GM (1,1) model established in this paper, we can predict the short-term changes of other data such as the change of population in the country or region, the change of economic growth and the prediction of Lanchester War.

The use of Exponential Growth Models can also predict long-term changes that, through these predictions, allow us to better predict the spread of some diseases and to exercise effective control prior to widespread disease. In addition, the investment company can also allocate funds reasonably according to the result predicted by the model in order to achieve the best return.

9 Memo

From: COMAP 81393

To: Chief Operating Officer

Date: 12 February 2018

Subject: Results and Recommendations

Dear Sir/Madam

We are the team you hire to predict the future language distribution and to choose a new office address. We wrote this memo to you to better explain the results our team predicted, as well as some suggestions.

First of all, our team first uses the two most popular and best-used forecasting methods: Exponential Growth Model and Gray Forecast Model. We have considered the national policy, economic and social pressures of various countries and the language(s) used in school. We have separately predicted the language distribution in the short time. Form the forecast result, we can see that Portuguese and Bengali have the most obvious growth trend as native language. Nevertheless, the largest number of languages used as native language are Chinese, Spanish and English while Japanese and German have a clear downward trend.

This is because the number of people who use English is growing year by year, while the natural growth rate of population in countries where English is the native language has risen. The government has adopted a policy of encouraging birth. The economic strength of these European countries has led more and more people learning English. The natural population growth rate in Japan and Germany has been declining year by year, resulting in a year-on-year decline in the number of people using Japanese and German. So, in the long run, there will be replaced results.

Of course, the above results do not take into account global migration and immigration. After our team reconsidered the impact of these factors, the current results were re-predicted and the results show that the number of people who use Chinese is still dominant and that Spanish will overtake English in the coming years.

The difference between the above results is that China has a large population base and a fast population growth rate, resulting in a large number of people using Chinese. The population of Spanish-speaking countries has increased in recent years and economic and military capabilities have also been steadily increasing. In contrast, English-speaking countries have seen negative growth in recent years, and will be overtaken by Spanish in the future.

These results are only a rough forecast of our current data, because of the incompleteness of available data, it may lead to the deviation of the forecast result from the actual situation. These factors can’t be avoided by our team, so your company can compare our forecast result as a reference, consider the actual situation of your company and make rational judgments.

Our team evaluated and ranked these cities by using the ideal solution. Considering the

annual amount of funds allocated to these cities, language distribution, accessibility and residents' happiness index, we can know the top 6 cities in the short term through the investigation of these factors and taking appropriate weights. In the long run, the top 6 cities are: Based on the above results, we suggest that your company establish new offices in the cities: Berlin, London, Paris, New Delhi, Moscow, Madrid.

In the longer term, the top 6 cities are: London, Madrid, Dhaka, Paris, New Delhi, Lisboa. Your company can refer to the short-term and long-term site selection, the specific decision of the final site.

As you can see, we gave the location of 6 new office addresses. However, our team analyzed a number of factors and made the following recommendations.

In the short term, your company can reduce the number of new offices to five locations: London, Paris, New Delhi, Moscow, Madrid. We give such a suggestion because we have found through surveys and the assessment of the forecast results that in the long run, Berlin has a negative growth trend and will eventually be replaced by Dhaka, so it will be eliminated from the original candidate list.

In the same way, our advice is only to provide your company with a reference. We hope your company can make the best decision based on its own actual conditions.

Yours sincerely,
COMAP 81393

References

[1] CIA.2018< https://www.cia.gov/> accessed 12 February 2018

[2] The Number of Population.2016< https://photius.com/> accessed 12 February 2018

[3] People.2018< http://bj.people.com.cn/> accessed 12 February 2018

[4] 360Library 2018< http://www.360doc.com/> accessed 12 February 2018

[5] Jun Lu, The Dilemma of Population in Contemporary Germany and the Guideline of Its Regulation and Control, 2009

[6] Guoqing Deng, Decreasing Birth Rates in Germany The UN Proposes to Relax Immigration Policies, 2001

[7] Weibing Peng, Study on the Historical Evolution and Influence Factors of Indian National Population Policy, 2014

[8] Qiuhong Shen, India's Population Development and Population Policy, 2013

[9] Dan Min, My opinion on the warming of small languages in foreign languages, 2010

[10] Changbao Mao, Comparison of the accuracy of the two population prediction models, 2009

[11] Jingli Jia, Two Types of Intuitionistic Fuzzy Sets Multi - attribute Decision Making and Improvement of Gray GM (1,1),2009

[12] Shoukui Si, Zhaoliang Sun, Mathematical Modeling Algorithms and Applications, Beijing: National Defense Industry Press, 2016.

[13] Ethnologue. 2018< http://www.ethnologue.com/ > accessed 10 February 2018.

Appendices

1. Exponential Growth Model (Take the example of how to use English as the native language)

```
 Clc,clear;
x=[1 6 10 16 18];
 y=[308.2 312.5 328.0 335.0 338.2];
 cftool    %Function is a*exp(b*x)-c
```

2. GM (1,1) (Take the example of how to use English as the native language)

```
 clc,clear
x0=[308.2 312.5 328.0 335.0 338.2]';
n=length(x0);
x1=cumsum(x0)
B=[-0.5*(x1(1:n-1)+x1(2:n)),ones(n-1,1)];
Y=x0(2:n);
u=B\Y
syms x(t)
x=dsolve(diff(x)+u(1)*x==u(2),x(0)==x0(1));
xt=vpa(x,5)
yuce1=subs(x,t,[0:n-1]);
yuce1=double(yuce1);
yuce=[x0(1),diff(yuce1)]
epsilon=x0'-yuce;
delta=abs(epsilon./x0');
```

3.Ideal Solution

```
 clc,clear
a=[ 2.57      6     9     6.725
      12.24    7     8     5.245
      1.31     6     5     6.361
      0.25     7     7     4.643
      2.44     7     6     4.404
      2.57     6     6     6.478
      1.47     6     7     5.856
      0.21     6     5     5.123
      3.65     6     6     6.994
      0.84     7     6     5.389
      ];
[m,n]=size(a);
for j=1:n
      b(:,j)=a(:,j)/norm(a(:,j));
end
w=[0.3 0.3 0.2 0.2];
```

```
c=b.*repmat(w,m,1);
Cstar=max(c);
C0=min(c);
for i=1:m
    Sstar(i)=norm(c(i,:)-Cstar);
    S0(i)=norm(c(i,:)-C0);
end
f=S0./(Sstar+S0)
[sf,ind]=sort(f,'descend')
```

4.Table 3

Language	Equation
Chinese	$119600e^{0.007332t}$ - 118700
Spanish	$17222e^{0.01855t}$ - 16900
English	$8960e^{0.03444t}$ - 8652
Hindi	$269300e^{0.0009628t}$ - 269100
Portuguese	$31511e^{0.005613t}$ - 31333
Bengali	$11100e^{0.01727t}$ -10922
Russian	$34600 - 34444e^{-0.00418t}$
Japanese	$42855 - 42733e^{-0.002868t}$
Javanese	$1419e^{0.05297t}$ - 1343.0
German	$2316 - 2221e^{-0.04365t}$
Korean	$1761e^{0.03737t}$ - 1694.0
French	$1838e^{0.03477t}$ - 1773.0
Telugu	$32400e^{0.002148t}$ - 32333.0
Marathi	$46277e^{0.001468t}$ - 46211.0
Turkish	$448e^{0.1145t}$ - 397.5
Tamil	$3159e^{0.02071t}$ - 3093.0

5.Table 4

Language / Weight	Chinese	Spanish	English	Hindi	Portuguese	Bengali	Russian	Japanese
Policy	0.75	0.45	0.5	0.75	0.4	0.55	0.35	0.55
Economic	0.8	0.35	0.7	0.65	0.5	0.4	0.55	0.65
Total	1.55	0.8	1.2	1.4	0.9	0.95	0.9	1.2
Language / Weight	Javanese	German	Korean	French	Telugu	Marathi	Turkish	Tamil
Policy	0.5	0.4	0.5	0.5	0.5	0.6	0.45	0.6
Economic	0.5	0.4	0.3	0.45	0.35	0.4	0.4	0.5
Total	1	0.8	0.8	0.95	0.85	1	0.85	1.1

6.Table 5

Language / Weight	Chinese	Spanish	English	Hindi	Portuguese	Bengali	Russian	Japanese
Policy	-0.2	0.2	0.1	-0.1	0.15	0.2	0.3	0.1
Economic	0.8	0.4	0.7	0.4	0.35	0.4	0.6	0.55
Social	0.4	0.2	0.2	0.3	0.2	0.3	0.15	0.25
School	0.5	0.3	0.3	0.3	0.3	0.25	0.25	0.35
Total	1.5	1.1	1.3	0.9	1	1.15	1.3	1.25
Language / Weight	Javanese	German	Korean	French	Telugu	Marathi	Turkish	Tamil
Policy		0.1		0.15	-0.1	-0.1	0.15	-0.1
Economic		0.4		0.4	0.4	0.4	0.35	0.4
Social		0.2		0.35	0.35	0.3	0.3	0.3
School		0.2		0.3	0.2	0.2	0.3	0.2
Total		0.9		1.2	0.85	0.8	1.1	0.8

7.Table 6

Language / Weight	Chinese	Spanish	English	Hindi	Portuguese	Bengali	Russian	Japanese
Policy	0.75	0.45	0.5	0.75	0.4	0.55	0.35	0.55
Economic	0.8	0.35	0.7	0.65	0.5	0.4	0.55	0.65
Total	1.55	0.8	1.2	1.4	0.9	0.95	0.9	1.2
Language / Weight	Javanese	German	Korean	French	Telugu	Marathi	Turkish	Tamil
Policy	0.5	0.4	0.5	0.5	0.5	0.6	0.45	0.6
Economic	0.5	0.4	0.3	0.45	0.35	0.4	0.4	0.5
Total	1	0.8	0.8	0.95	0.85	1	0.85	1.1

8. Table 7

Language / Weight	Chinese	Spanish	English	Hindi	Portuguese	Bengali	Russian	Japanese
Policy	-0.2	0.2	0.1	-0.1	0.15	0.2	0.3	0.1
Economic	0.8	0.4	0.7	0.4	0.35	0.4	0.6	0.55
Social	0.4	0.2	0.2	0.3	0.2	0.3	0.15	0.25
School	0.5	0.3	0.3	0.3	0.3	0.25	0.25	0.35
Total	1.5	1.1	1.3	0.9	1	1.15	1.3	1.25
Language / Weight	Javanese	German	Korean	French	Telugu	Marathi	Turkish	Tamil
Policy		0.1		0.15	-0.1	-0.1	0.15	-0.1
Economic		0.4		0.4	0.4	0.4	0.35	0.4
Social		0.2		0.35	0.35	0.3	0.3	0.3
School		0.2		0.3	0.2	0.2	0.3	0.2
Total		0.9		1.2	0.85	0.8	1.1	0.8

9. Table 8

Language	Country	Rate of Immigration (migrants/1000 population) and Total Population (million)							
		2012	Total	2013	Total	2015	Total	2016	Total
Chinese, Mandarin	China	-0.34	1330.14	-0.33	1343.24	-0.44	1349.59	-0.32	1355.69
Spanish	Spain	2.73	46.51	5.02	47.04	8.31	47.37	7.24	47.74
	Argentina	0	41.34	0	42.19	0	42.61	0	43.02
English	United Kingdom	2.61	62.35	2.59	63.05	1	63.40	2.56	63.74
	United States	4.25	310.23	3.62	313.85	2.43	316.67	2.45	318.89
	Canada	5.64	33.76	5.65	34.30	5.66	34.57	5.66	34.83
	Australia	6.13	21.52	5.93	22.02	5.65	22.26	5.74	22.51
Hindi	India	-0.05	1173.11	-0.05	1205.07	-0.04	1220.80	-0.05	1236.34
Portuguese	Portugal	3.06	10.74	2.9	10.78	2.67	10.80	2.74	10.81
Bengali	Bangladesh	-2.12	156.12	-1.04	161.08	0.46	163.65	-0.02	166.28
Russian	Russian Federation	0.28	139.39	0.29	142.52	1.69	142.50	1.69	142.47
Japanese	Japan	0	126.80	0	127.37	0	127.25	0	127.10
Javanese	Indonesia	-1.23	242.97	-1.08	248.65	-1.16	251.16	-1.18	253.61
German, Standard	Germany	2.19	82.28	0.71	127.37	1.09	81.15	1.06	81.00
Korean	South Korea	0	48.64	0	48.86	2.6	48.96	0	49.04
French	France	1.47	64.77	1.1	65.63	1.24	65.95	1.09	66.26
Telugu	India	-0.05	1173.11	-0.05	1205.07	-0.04	1220.80	-0.05	1236.34
Marathi	India	-0.05	1173.11	-0.05	1205.07	-0.04	1220.80	-0.05	1236.34
Turkish	Turkey	0.53	77.80	0.5	79.75	2.16	80.69	0.46	81.62
Tamil	India	-0.05	1173.11	-0.05	1205.07	-0.04	1220.80	-0.05	1236.34

10.Table 10

City	GDP(tillion)	Language	Traffic	Happiness Index
London	$2.57	6	9	6.725
Shanghai	$12.24	7	8	5.245
Madrid	$1.31	6	5	6.361
Dhaka	$0.25	7	7	4.643
NEW DELHI	$2.44	7	6	4.404
Paris	$2.57	6	6	6.478
Moscow	$1.47	6	7	5.856
Lisboa	$0.21	6	5	5.123
Berlin	$3.65	6	6	6.994
Ankara	$0.84	7	6	5.389