

For office use only
T1 _____
T2 _____
T3 _____
T4 _____

Team Control Number

78431

Problem Chosen

B

For office use only
F1 _____
F2 _____
F3 _____
F4 _____

2018
MCM/ICM
Summary Sheet

The blueprint of our future world's languages - A language population dynamics model

Summary

We propose a language population dynamics model to investigate the trends of world's language. We work out a primary language population dynamics model first and then use comprehensive evaluation to make amendments to this model.

First, we reasonably put emphasis on the main influences of language population caused by the fertility, mortality and migration conditions. Then, we work out the primary language population dynamics model.

Next, we add other factors as modulation terms to the primary model. We use method for comprehensive evaluation to achieve this goal, TOPSIS specifically. We transform this valuation of language's else factors into the increase in population. Then, we work out the revised language population dynamics model as our final model.

Afterwards, we draw out the world's map in the unit of language. Finally, we pin the suitable international offices on this world's map of language. We continue to use TOPSIS and add more business qualified factors like English popularity and economic conditions to the model.

Given the result of our model, we are presented with the question of where suitable international offices should be located such that our international offices have optimal quality of employees. We present the curve of 12 main languages' population over time both in narrow sense and in generalized sense. We also present the language value world map. We advised to set our offices in New Delhi, India; Ciudad de MÃ½xicow, Mexico; Dhaka, Bangladesh; Jakarta, Indonesia; Cairo, Egypt; Rio de Janeiro, Brazil.

Keywords:

Leslie matrix; population prediction; TOPSIS

Contents

1	Introduction	2
1.1	Background	2
1.2	Restatement of the Problem	2
1.3	Our approach	2
2	Assumptions and Justifications	3
3	Notation	4
4	Primary model for language population dynamics	4
4.1	Model Overview	4
4.2	Concrete Model	5
4.2.1	Choosing suitable representative countries	5
4.2.2	Forming the model for population dynamics	6
5	Model for evaluation of other language factors as amendment	7
5.1	Model Overview	7
5.2	Concrete model	8
6	Model analysis	10
6.1	Investigate trends of global languages	10
6.1.1	The numbers of native speakers and total language speakers over time	10
6.1.2	The rank change of 10 main language	11
6.1.3	The geographic distributions of language	12
6.2	Location options for new offices	12
6.3	Sensitivity analysis	14
7	Conclusion	15
7.1	Strength and weakness	15
7.1.1	Strength	15
7.1.2	Weakness	16

1 Introduction

1.1 Background

Of the some 6900 languages in the world, the native tongue of most of world's population concentrate in only a few, making speakers of those languages populous. With globalization in process, chances are that more people will speak one or more additional languages. Population speaking a language may fluctuate due to a variety of factors. The natural population change of the native speakers of a language, migration and government policies are some of the main factors while other small factors play a non-negligible role.

1.2 Restatement of the Problem

We chose the problem of investigating trends of global languages and location of new international office.

We are offered with *List of languages by total number of speakers*¹. In the list, we get the population of L2 speakers.

However, there are two notions of second language (or L2) that either

- A language that is not the native language of the speaker, but that is used in the locale of that person. In contrast, a foreign language is a language that is learned in an area where that language is not generally spoken by the community as a whole.
- Any language learned in addition to one's native language, especially in the context of second-language acquisition.

We focus exclusively on the first definition. This definition agrees with accessible data more closely because the formal data bank of the United Nations defines second language as the first definition. For example, English is regarded as second language in Hong Kong, China, while a considerable share of the population of English speakers in China mainland is not counted.

1.3 Our approach

We are aimed at modeling the population mobility of the language speakers.

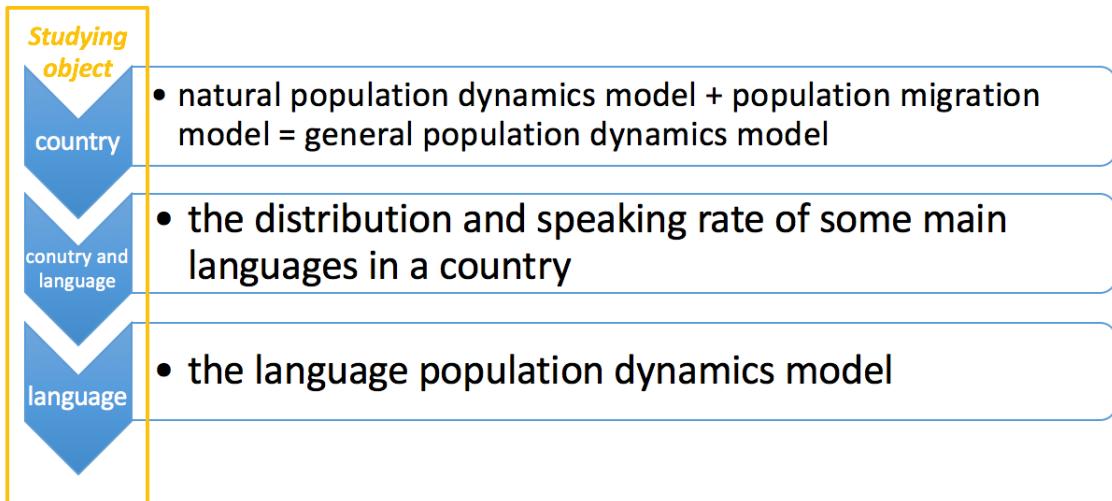
First, we reasonably put emphasis on the main influences of language population caused by the fertility, mortality and migration conditions. Then, we work out the primary language population dynamics model.

Next, we add other factors as modulation terms to the primary model. For example, culture factors like booming tourism and culture influence of a language, the linguistic factor like language family, etc. We use method for comprehensive evaluation to achieve

¹Retrieved from https://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers on January 17, 2018.

this goal, TOPSIS specifically. Then, we work out the revised language population dynamics model as our final model.

Finally, we can pin the suitable international offices. We continue to use method for comprehensive evaluation to achieve this goal. The determinant includes the English popularity and economic conditions besides the final language population dynamics model, for the aim of locating the appropriate offices with easier recruitment of multilingual employees and higher business value.



2 Assumptions and Justifications

1. **Age-specific death rate remains constant over time.**
2. **Age-specific fertility rate of female remains constant over time.**
3. **The sex ratio from the age of 15 to the age of 50 remains constant.** This age range is regarded as the child-bearing period of female and we use the existed data to work out our model.
4. **Only natural change of population be considered.** The change of fertility, mortality and migration rate of the population because of local wars, natural disasters or any other special reasons should not be taken into consideration.
5. **Language policy change little.** The transform in the language policy of a country and the language(s) used in schools.
6. **The descendants of immigrants can speak the language(s) of their adopted countries.** We consider the descendants have assimilated.
7. **The age structure of immigrants remains constant.**
8. **The growth rate of the major language(s) in a country approximately equals that of country's population.** We suppose the living conditions of different regions in one same country are similar. The overall growth rate can represent the individual overall growth.

9. The ratio of the representative country(s)' population which speak a certain language to its total population of speakers in all related countries remains constant. The major language(s) in the country has(have) a considerable population of speakers. The reason of this assumption is similar to the prior assumption. The indicators of main countries' population can represent that of all related countries.

$$\frac{N_1}{N_2} \equiv C \quad (1)$$

In this equation, N_1 is the number of a language's speakers in representative countries and N_2 is the total number of a language's speakers in all related countries

3 Notation

Symbol	Definition
$\Psi(t)$	Number of newborn in a period of time from t to $t + 1$
$\beta(t)$	Total fertility rate
$x_i(t)$	Number of people at the age of i and $i + 1$ in a period of time from t to $t + 1$
μ_{00}	Infant mortality rate
$\mu_i(t)$	Age-specific death rate
$g_i(t)$	Number of net migration people at the age of i and $i + 1$ in a period of time from t to $t + 1$

4 Primary model for language population dynamics

4.1 Model Overview

We generally study the population change in units of country. Therefore, we should first refine our modelling work to a country scope. To simplify the modelling works, we select the suitable countries as studying objects according to the lists of top ten languages given. We choose about 20 appropriate countries.

This primary language population dynamics model is influenced by the fertility, mortality and migration conditions of these selected countries. Therefore, we design our working process as follow:

Step 1: Work out the natural population dynamics model and population migration model of a *country*.

Step 2: Involve the studying objects of *language* with the help of data concerning the distribution and speaking rate of some main *languages* in a certain *country*.

Step 3: Change the studying object from *country* to *language* for the ultimate purpose of investigating the trend of global languages. Then we form the language population dynamics model.

4.2 Concrete Model

4.2.1 Choosing suitable representative countries

Above all, the usage of major languages is greatly influenced by the countries with large population and our selection put more weight on the population of the countries speaking a certain language. The list of the countries with the greatest population are listed in the table and the major languages spoken in these countries are specified. From the table, we can predict that as the population of the countries go down, the influence of the country towards the languages will become weaker and weaker. As a result, the countries with the twenty largest population are selected.

Rank	Country Name	Population (2016 statistics)	Percentage of world population	Accumulated percentage
1	China	1378665000	18.5%	18.5%
2	India	1324171354	17.7%	36.2%
3	United States	323127513	4.3%	40.5%
4	Indonesia	261115456	3.5%	44.0%
5	Brazil	207652865	2.8%	46.8%
6	Pakistan	193203476	2.6%	49.4%
7	Bangladesh	162951560	2.2%	51.6%
8	Russian Federation	144342396	1.9%	53.5%
9	Mexico	127540423	1.7%	55.2%
10	Japan	126994511	1.7%	56.9%
11	Egypt, Arab Rep.	95688681	1.3%	58.2%
12	Congo, Dem. Rep.	78736153	1.1%	59.3%
13	France	66896109	0.9%	60.1%
14	United Kingdom	65637239	0.9%	61.0%
15	Spain	46443959	0.6%	61.6%
16	Canada	36286425	0.5%	62.1%
17	Morocco	35276786	0.5%	62.6%
18	Saudi Arabia	32275687	0.4%	63.0%
19	Malaysia	31187265	0.4%	63.5%
20	Portugal	10324611	0.1%	63.6%

More specifically, we can see that the 20th country count for only 0.1% of the world population, indicating that the countries below it have even smaller population proportions, making its influence on the world language change even weaker. From another perspective, the top twenty countries account for 63.6% of the world population. Therefore, they can well represent the trend of the main population change in the world.

CHINESE (1090)				Bengali (268)			
CHINA	965.3	91.10%		India	117.3		
				Bangladesh	150		
English (982)				Total	267.3	100%	
India	161			Portuguese (229)			
US	258.4			Brazil	202.9		
UK	66			Portugal	10		
Canada	23.4			Total	212.9	93%	
Total	508.8	51.90%		French (229)			
Spanish (527)				Congo	45.1		
US	42			France	67		
Mexico	126.7			Canada	36		
Spain	46			Total	148.1	64.70%	
Total	214.7	40.70%		Punjabi (148)			
Arabic (422)				Pakistan	86.9	58.70%	
Egypt	65.3			Japanese (129)			
Morocco	29.8			Japan	129		
Saudi Arabia	28.8			Hindi (380)			
Total	123.9	29.40%		India	378	99%	
Malay (281)				Russia	267		
Indonesia	208.8						
Malaysia	31						
Total	239.8	85.30%					
Russian (267)							
Russia	267						

We get this data through the statistics of representative speaking countries of a certain language.

According to Leslie model, we can conclude the following equations:

$$\Psi(x) = \beta(t)k \sum_{15}^{50} x_i(t) \quad (2)$$

$$x_0(t) = (1 - \mu_{00})\Psi(t) + g_0(t) \quad (3)$$

$$x_1(t+1) = (1 - \mu_0(t))x_0(t) + g_1(t) \quad (4)$$

$$x_m(t+1) = (1 - \mu_m - 1(t))x_m - 1(t) + g_m(t) \quad (5)$$

$$N(t) = \sum_0^m x_i(t) \quad (6)$$

4.2.2 Forming the model for population dynamics

According to the third assumption in the second section, k , $u_i(t)$, $x_i(2010)$ and $g_i(t)$ are accessible from the database of the United Nations. We can calculate $N(t)$ from the statistics above.

	China	USA	UK	Canada	India	Mexico	Egypt	Morocco	Saudi Arabia	Indonesia	Malaysia	Russia	Bangladesh	Brazil	Portugal	Congo	France	Japan
2015	1415	328	67	36	1456	122	83	33	30	262	30	155	157	205	1	4	67	128
2020	1444	332	67	36	1504	128	91	35	33	277	32	152	166	211	1	5	67	126
2025	1461	336	67	36	1532	135	99	37	35	292	33	149	174	217	1	6	67	124
2030	1468	339	67	36	1550	140	108	40	36	305	34	146	183	220	1	7	68	122
2035	1464	340	67	35	1557	144	115	41	38	316	35	143	189	222	1	8	68	118
2040	1445	339	67	35	1549	148	123	43	38	325	36	138	195	223	1	9	67	115
2045	1414	337	66	34	1530	150	129	44	39	332	36	134	199	222	1	9	67	112
2050	1376	334	66	33	1506	152	135	45	39	336	36	130	201	219	1	10	66	109
2055	1332	331	65	32	1478	152	141	46	38	339	36	126	202	215	1	12	66	106
2060	1284	327	64	31	1449	152	146	47	38	339	36	122	201	211	1	13	65	103
2065	1236	324	63	30	1416	151	150	47	36	338	36	118	199	205	1	14	64	100
2070	1189	321	62	29	1381	150	153	47	35	336	35	114	196	199	1	15	64	96
2075	1146	317	61	28	1349	147	156	47	34	333	34	111	192	192	1	16	64	93
2080	1105	314	61	27	1317	145	157	46	32	329	33	108	187	185	1	17	63	91
2085	1068	310	60	26	1287	142	159	46	31	324	32	105	181	179	1	18	63	89

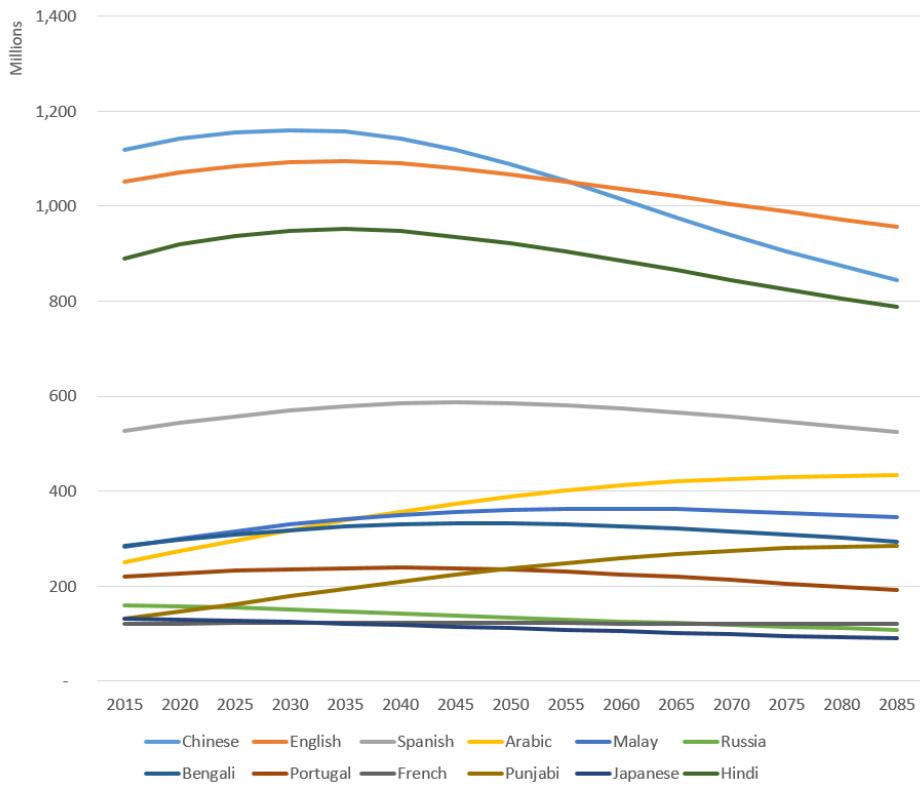
Population prediction of the countries

Besides, given the assumptions 8 and 9, the predictions of the country can be converted to the predictions of the languages.

The primary language population dynamics model

	Chinese	English	Spanish	Arabic	Malay	Russia	Bengali	Portugal	French	Punjabi	Japanese	Hindi
2015	1,119	1,052	527	250	282	160	285	220	120	132	131	890
2020	1,141	1,070	543	274	299	157	297	227	121	147	130	920
2025	1,154	1,083	558	296	315	154	308	232	122	163	127	936
2030	1,160	1,092	570	318	329	151	318	236	123	178	125	948
2035	1,157	1,095	579	338	341	147	325	238	123	194	121	952
2040	1,142	1,090	584	357	350	143	330	238	124	209	118	947
2045	1,118	1,080	586	373	356	138	332	237	123	223	115	936
2050	1,088	1,067	585	388	361	134	332	235	123	236	112	921
2055	1,053	1,052	581	400	363	130	330	230	122	248	108	904
2060	1,015	1,036	575	411	363	126	327	225	122	258	105	886
2065	977	1,020	566	420	362	122	322	219	121	267	102	866
2070	940	1,004	556	426	359	118	316	213	121	274	99	845
2075	906	988	546	430	355	114	309	206	121	279	96	825
2080	873	972	535	433	350	111	301	198	121	283	93	805
2085	844	956	523	434	344	108	293	191	121	285	91	787

The primary language population dynamics model



5 Model for evaluation of other language factors as amendment

5.1 Model Overview

Except for the main factors studied above, we should also take other factors into consideration. Here, we list four factors:

- **Economic condition.** Economic condition tells the activeness of a country's international business. The better the economic condition, the more attractive the foreign investments.

- **Potential second language speakers.** According to the definition of L2 mentioned in the restatement, the property of the language may change over time. For example, English is regarded as foreign language at present in Shanghai, China. However, according to the EF EPI², the English proficiency index of Shanghai is 56.76 while that of Hong Kong is 55.81. At present, English has become the second language of Hong Kong according to the Ethnologue³. We can reasonably assume that in the near future, English will be the second language of people in Shanghai. These citizens of Shanghai are regarded as potential second language speakers.
- **Cultural factors including global tourism and culture influence.** Global tourism promotes language learning. What's more, American TV dramas, British TV dramas and Japanese animation are just some representative of the cultural influence nowadays. These cultural factors are non-negligible.
- **Linguistic factor of language family.** Some language families may have priorities in terms of difficulty of acquisition due to the history of language development. For example, Indo-European languages are the most widely distributed language in the contemporary world. Today, about 46% of the human population speaks an Indo-European language as a first language, by far the highest of any language family. This promises Indo-European languages to have priority over other language families.

5.2 Concrete model

We use TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution)⁴ to evaluate these four factors.

$$E = \begin{bmatrix} 1 & 4 & 2 & 8 \\ 1/4 & 1 & 1/2 & 2 \\ 1/2 & 2 & 1 & 4 \\ 1/8 & 1/2 & 1/4 & 1 \end{bmatrix}$$

Pairwise comparison matrix

²The English proficiency index research studied by the EF company <https://liuxue.ef.com.cn/epi/regions/asia/china/>

³Retrieved from <https://www.ethnologue.com/>.

⁴ <https://en.wikipedia.org/wiki/TOPSIS>

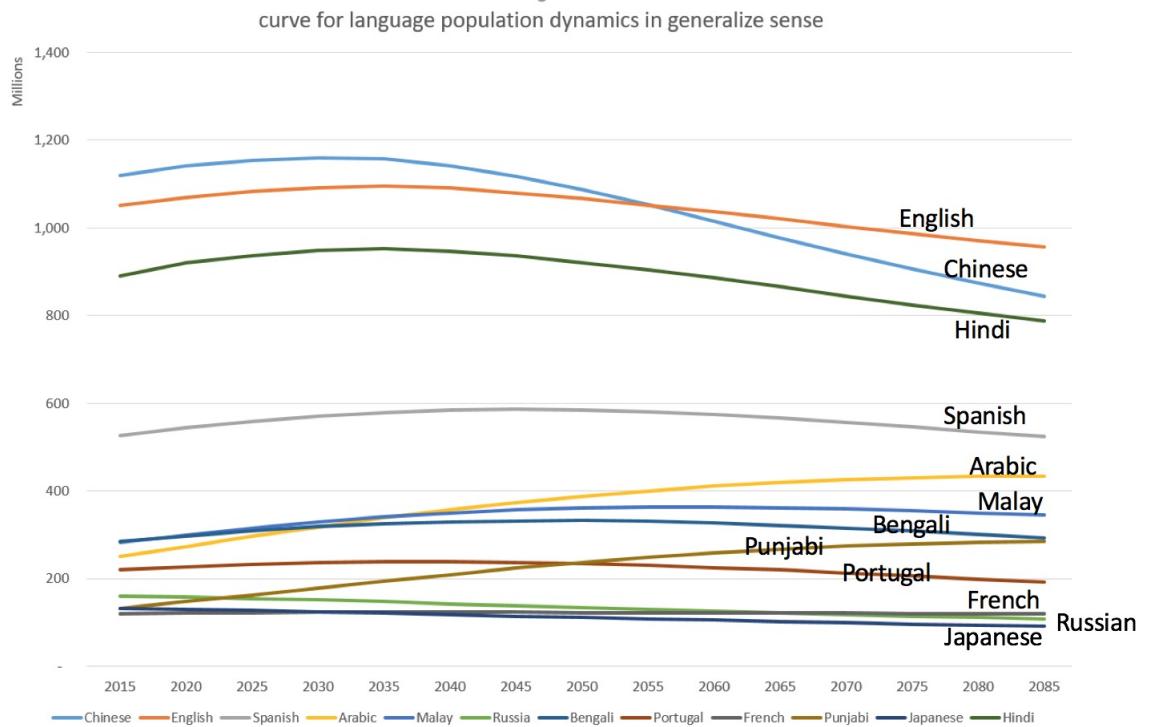
	Economic condition	Cultural factors	Potential speakers	Linguistic factor
Chinsese	A	B	A	B
English	A	A	A	A
Spanish	B	A	B	A
Arabic	B	C	B	B
Malay	B	C	B	B
Russian	A	B	B	A
Bengali	C	C	C	A
Potugal	B	B	B	A
French	B	A	A	A
Punjabi	C	C	C	A
Japanese	A	B	A	B
Hindi	A	C	C	A

The evaluation of language

country	value	Rank
Chinese	0.750	2
English	1.000	1
Spanish	0.471	11
Arabic	0.286	6
Malay	0.286	12
Russia	0.734	9
Bengali	0.083	3
Portugal	0.341	8
French	0.495	4
Punjabi	0.083	5
Japanese	0.750	7
Hindi	0.644	10

The blank of value determines the rank of each language. Then we use this rank to make some amendments to the primary language population dynamics model. According to the data of *Languages for the Future*⁵, we assume that these factors will at most influence 3% of the language population. The change rate of language population can follow the linear change in the range of 0 to 3%. Then we get the curves of language population dynamics model in generalized sense.

⁵The model is studied by the British Council, www.britishcouncil.org



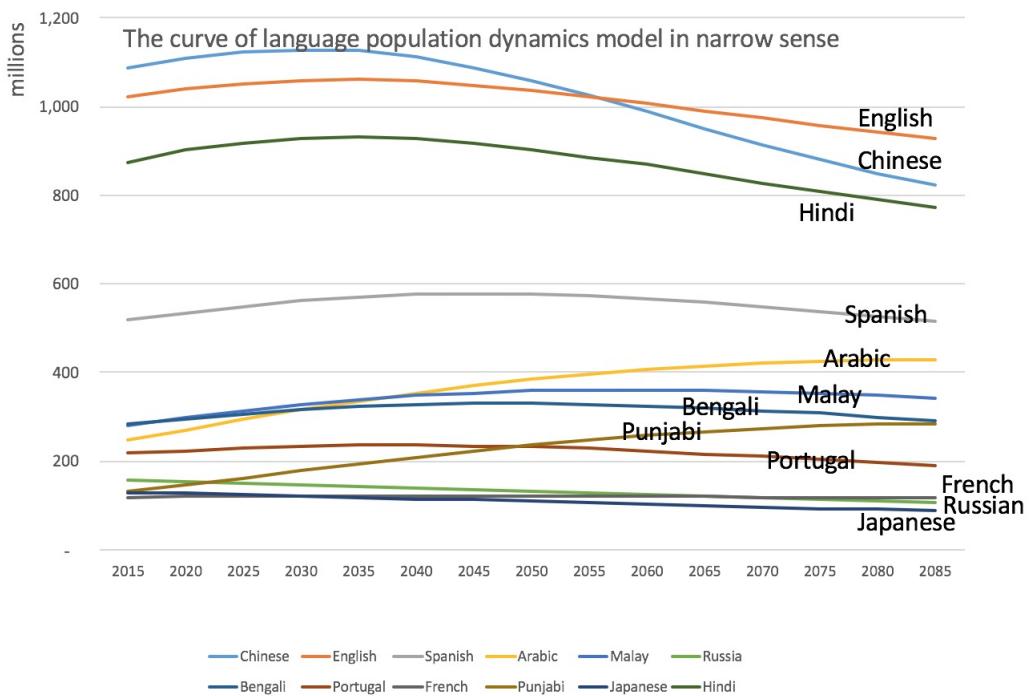
6 Model analysis

6.1 Investigate trends of global languages

6.1.1 The numbers of native speakers and total language speakers over time

We get the curve of 12 main languages' population over time both in narrow sense and in generalized sense. However, this curve describes the property of total language speakers. In order to investigate the numbers of native speakers in the future, we just need to use the curve of narrow sense because the mobility rate of the population of native speakers is only related to the natural mobility of the population. Other factors give more influence to the number of second language speakers.

	2070
English	974,618,746
Chinese	914,726,856
Hindi	828,050,011
Spanish	548,052,074
Arabic	421,828,043
Malay	356,091,947
Bengali	314,245,795
Punjabi	273,448,199
Portuguese	210,039,149
French	118,853,047
Russian	115,418,416
Japanese	96,369,252



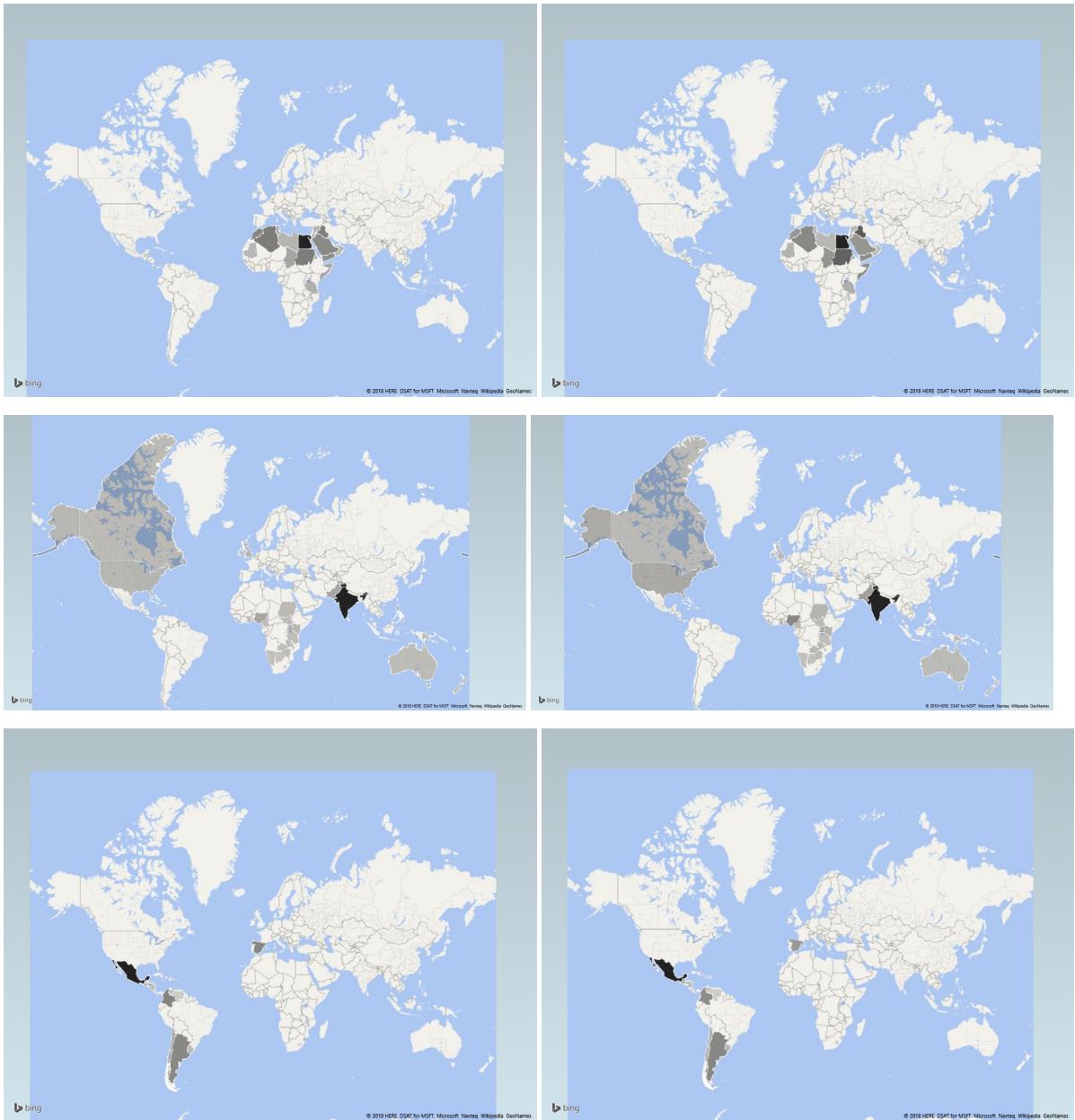
6.1.2 The rank change of 10 main language

	2015		2070
Chinese	1,118,567,546	English	1003857308
English	1,052,088,529	Chinese	939,881,845
Hindi	890,045,337.3	Hindi	844,611,010.8
Spanish	526,834,905.8	Spanish	556,272,855.5
Bengali	284,671,366.4	Arabic	426,046,323.2
Malay	282,420,672.6	Malay	358,762,636.6
Arabic	250,011,399.3	Bengali	315,817,024.3
Portuguese	219,879,681.9	Punjabi	274,131,819.7
Russian	159,957,533.6	Portugal	212,664,638.4
Punjabi	131,665,152.7	French	120,932,974.9
Japanese	131,174,333	Russian	118,015,329.9
French	120,395,714	Japanese	98,778,483.3

We can find the rank of top ten main language change. French become a ten main language and Russian will be replaced.

6.1.3 The geographic distributions of language

We selected three representative languages, Arabic, English and Spanish to do the geographical distribution of the analysis.

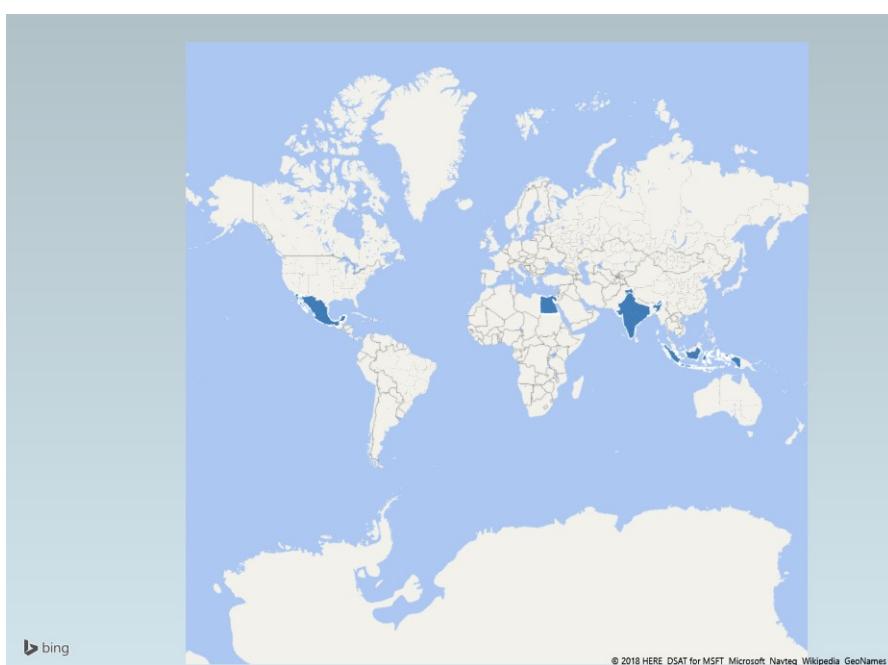


6.2 Location options for new offices

In the suppose of our problem, we have had two offices both in New York City in the United States and Shanghai in China. Therefore, English and Chinese can be neglected in this analysis. Then we choose Hindi, Spanish, Bengali, Malay, Arabic and Portugal as the working language in the office besides the basic requirements of English acquisition. In order to have more business value, we choose the relatively developed region among the entire area which speaks the language. The location of the office should be New Delhi, India; Ciudad de Mexicowe, Mexico; Dhaka, Bangladesh; Jakarta, Indonesia; Cairo, Egypt; Rio de Janeiro, Brazil.

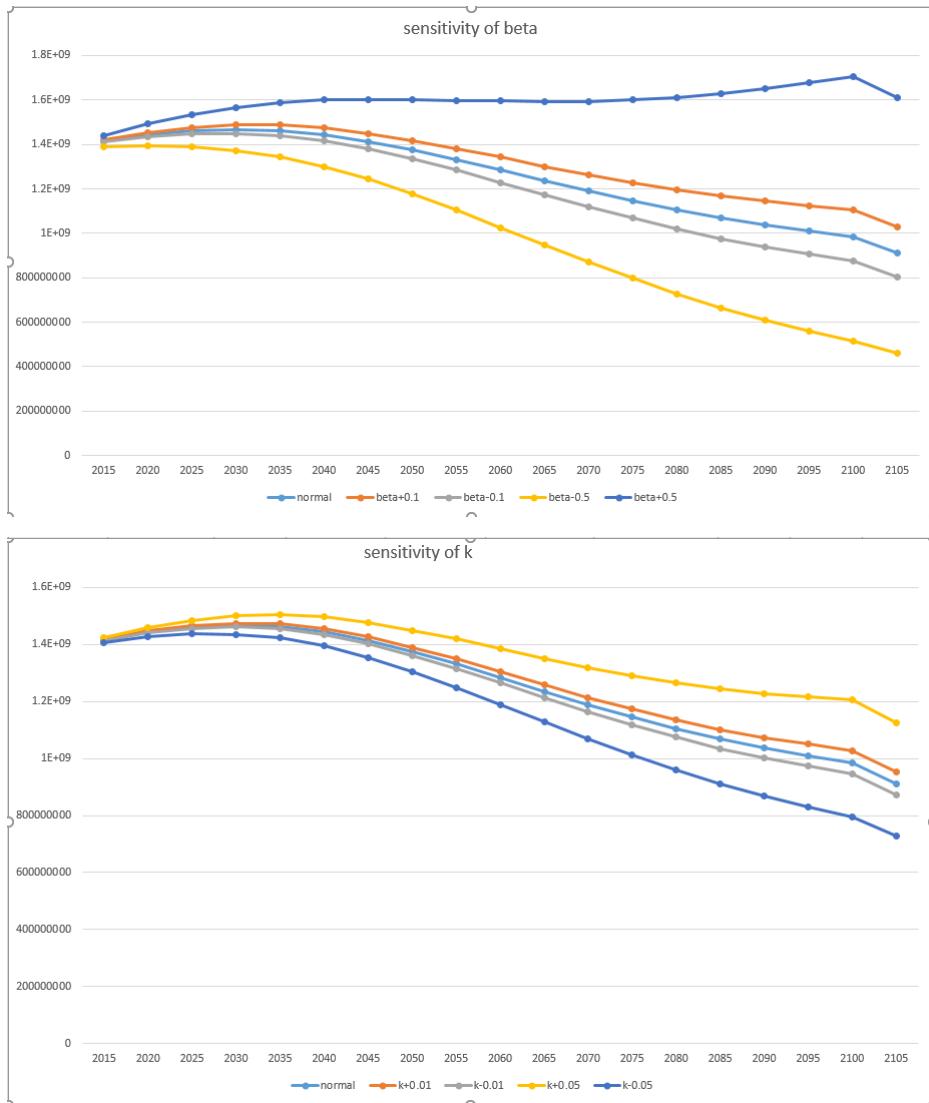
What's more, we suggest that the company only open four international offices for long-term development in New Delhi, India; Ciudad de Mexicowe, Mexico; Jakarta, In-

donesia; Cairo, Egypt.



6.3 Sensitivity analysis

Sensitivity Analysis: We choose China to do a sensitivity analysis of beta and k, the results are as follows:



Both β and k have an impact on the model, with beta can even change the upward or downward trend.

7 Conclusion

7.1 Strength and weakness

7.1.1 Strength

Various factors have been taken into account, of which the population model takes into account the age structure and population migration, the projected population is consistent with reality, other factors are taken into account, and the flexibility of the model is enhanced.

7.1.2 Weakness

Fertility mode, the sex ratio and the impact of the model on the model is relatively large, the impact of other factors on the model is not quantitative enough.

Memo

To: Chief Operating Officer From: A MCM Team Date: February 13, 2018 Subject: Investigation of global language trends and location options for new offices.

Mr. Officer,

We have attached the report to this email, but we also wanted to be given the honor of quickly discussing the trend noticed and our results of suitable location for new offices.

We've investigated 12 world's main language: Chinese, English, Spanish, Arabic, Malay, Russia, Bengali, Portugal, French, Punjabi, Japanese and Hindi.

In the process, we reasonably put emphasis on the main influences of language population caused by the fertility, mortality and migration conditions first.

Then, we work out the primary language population dynamics model.

Next, we add other factors as modulation terms to the primary model. For example, culture factors like booming tourism and culture influence of a language, the linguistic factor like language family, etc. We use method for comprehensive evaluation to achieve this goal, TOPSIS specifically. Then, we work out the revised language population dynamics model as our final model.

Finally, we can pin the suitable international offices. We continue to use method for comprehensive evaluation to achieve this goal. The determinant includes the English popularity and economic conditions besides the final language population dynamics model, for the aim of locating the appropriate offices with easier recruitment of multilingual employees and higher business value.

We have had two offices both in New York City in the United States and Shanghai in China. Therefore, English and Chinese can be neglected in this analysis. Then we choose Hindi, Spanish, Bengali, Malay, Arabic and Portuguese as the working language in the office besides the basic requirements of English acquisition. In order to have more business value, The location of the office should be New Delhi, India; Ciudad de Mexico, Mexico; Dhaka, Bangladesh; Jakarta, Indonesia; Cairo, Egypt; Rio de Janeiro, Brazil.

What's more, we suggest that the company open four international offices for long-term development in New Delhi, India; Ciudad de Mexico, Mexico; Jakarta, Indonesia; Cairo, Egypt;

Please let me know if you have any questions.

Best wishes

MCM Team

References

- [1] Wikipedia, the free encyclopedia. List of languages by total number of speakers, retrieved from https://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers on February 13, 2018.
- [2] EF Education First. The English proficiency index research, retrieved from <https://liuxue.ef.com/epi/regions/asia/china/> on February 13, 2018.
- [3] Simons, Gary F. and Charles D. Fennig (eds.). 2017. Ethnologue: Languages of the World, Twentieth edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>.
- [4] Wikipedia, the free encyclopedia. TOPSIS, retrieved from <https://en.wikipedia.org/wiki/TOPSIS> on February 13, 2018.
- [5] United Nations Statistics Division, Population by age, sex and urban/rural residence, retrieved from <http://data.un.org/Data.aspx>
- [6] Jiang Qiyuan. Mathematical Experiment and Mathematical Modeling [J]. Mathematics and Practice, 2001, 31 (5): 613-617.
- [7] Si Shoukui, Sun Xi Jing Mathematical modeling algorithm and application [M]. National Defense Industry Press, 2011.