For office use only	Team Control Number	For office use only
T1 _____	**82277**	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	**B**	F4 _____

2018

MCM/ICM

Summary Sheet

Let there be Language

Summary

"Let there be light." - an English translation from Hebrew Bible, became well-known to the world by the wide spread of English. Language plays an important role in culture, religion and trade. To investigate the trends of languages and locations for new offices of a multinational company, we build a prediction model and a profit model in this paper.

The first model provides two predictions for total language speakers and language distribution respectively. On one hand, we use Logistic Regressive to simulate population growth for determining the number of native speakers and Analytic Hierarchy Process to quantify the trend of newly-born population choosing second languages based on four factors. Concluding that English will exceed Chinese to be the No.1 language 50 years later. On the other hand, we simulate language distributions in several countries based on the data of immigrant, emigrant and population growth. The result is that if we do not limit immigrants, the proportion of native speakers in a country with low birth rate may decrease by time.

The second model discusses the relation between language distribution and office profit. We use work efficiency to connect the two parameters and determine office language. We conclude that using local language as the one being used in office gains more profit. Japan, German, France, Britain, India and Russia are the best locations for new offices. Russia can be removed from the list if necessary, due to providing smaller profit compared with the other five countries.

Contents

1 Introduction

1.1 Background

As the way of communication for human society, various languages were created in different areas because of the inconvenience of transportation and communication in ancient times. Globalization has become an inevitable trend with the widespread use of computers and networking. As a result, distant languages conflict when disseminating information at exceedingly high speeds. Meanwhile, the competition between them is getting more and more intense. Some languages fell out of use or even became extinct in this process as new languages and vocabulary are developed. Currently, except their native languages, people tend to master a second language in order to keep the communication and connection with the rest of the world. English, which has been used as a second language by largest number of people, is typically regarded as a significant one among them.

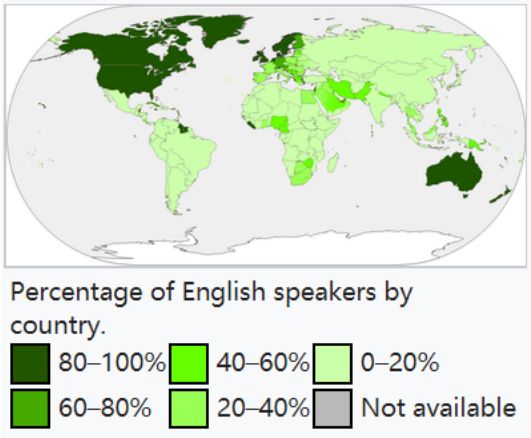


Figure 1: distribution of English speakers [1]

As English becoming the leading language of international discourse, the number of multinational corporations is increasing over time. Usually the employees in these company are requested to master English, which helps expand business across countries and achieve a dominant position in the world market. There is no doubt that communicating with the same language adeptly could improve the efficiency and therefore benefits the corporations.

1.2 Restatement of the Problem

We are required to build a model to predict the population and geographical distribution of languages. Based on this model, we are also required to decide where to open

international offices for most profits. Therefore, the problem can be divided into three parts:

- Build a model to predict the total number of language speakers and the geographical distribution in the future.
- Build a generic model to explain the relation between profit and language in an office and provide a staff assignment strategy.
- Use the prediction model and the profit model to decide the best locations of offices.

2 Symbols

Symbols	Meanings
S	total number of speakers of a language
P	population
r	growth rate of population
K	carrying capacity
t	time
P_{rf}	profit
R	revenue
C	cost
n	number of staffs in an office
E	work efficiency

3 Model I Time-depend Prediction Model

3.1 Overview

The number of speakers of a language and the languages' distribution in geography are two problems for us to solve. Both of them are related to several factors with low correlation. Thus, we build two time-depend prediction model respectively.

3.2 Assumptions

- Only consider top 20 languages in the ranking. Because the speaker numbers of the rest languages are too small.
- Ignore the countries with small populations. Some countries have such small populations that have no obvious effect on our calculate.
- One's native language is decided by one's birthplace.

- All new populations in the world decide their second language since their births.
- No one master a third language. The number of persons who can speak three language is too small. Besides, it is hard for us to find a credible data.
- One's second language will not change.
- No dialects. We just consider the official language of a country. The dialect speakers are combined into official language speakers.

3.3 Establishment

3.3.1 Language Speakers

The speakers of a language S can be divided into to parts: native speakers(use language as L1) S_{L1} and second language speakers(use language as L2) S_{L2}. The relation between them is

$$S = S_{L1} + S_{L2} \tag{1}$$

(1)Native Speakers

Since one's L1 is decided by one's birthplace, so S_{L1} is equal to the population of countries whose government claims this language as official language. We use Logistic Regressive to predict the population a country. The prediction equation [3] is

$$P(t) = \frac{K}{1 + \frac{K-P_0}{P_0}\, e^{-rt}} \tag{2}$$

where:

$P(t)$	population of tth year
P_0	initial population
t	time
r	constant,represent growth rate
K	carrying capacity

Based on the historical population data of a country, K and r can be measured by MATLAB [4]. Then a prediction of population growth is available.

(2)Second language speakers

Now globalization has been a trend for some years. We consider that transmit infor-

mation on website is easy today and even easier in the future. One whether choose a language as one's L2 is influenced by four factors:

- GDP of countries that use the language
- Population of the language
- Degree of difficulty
- Extent of transmission on website

To quantify the trend of learning a language, we define **Influence of Language**(IL) here. IL is a probability for new population in a country choosing their L2. To measure IL, we use Analytic Hierarchy Process(AHP) [5] to weigh the 20 languages by the four factors. Then IL can be described as

$$IL = \frac{w_{Language}}{w_{AllLanguage}} \tag{3}$$

where:

$w_{Language}$ language's weigh determined by AHP

$w_{AllLanguage}$ sum of all the languages' weighs

In AHP, the relatively importance between criteria is presented by number 1-9. Then a criteria Judging matrix A can be made. Then eigenvector μ and maximum eigenvalue λ_{max} is calculated by MATLAB. After consistency check, which uses coincidence indicator table, A can be determined.

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.52	0.89	1.11	1.25	1.35	1.40	1.45	1.49
n	11	12	13	14	15	16	17	18	19	20
RI	1.52	1.54	1.56	1.58	1.59	1.5943	1.6064	1.6133	1.6207	1.6292

Table 1: coincidence indicator table [6]

Then we determined the matrix A, whose weight vector is $\mu = (0.6053, 0.0670, 0.0655, 0.2621)$.

$$A = \begin{bmatrix} 1 & 7 & 9 & 3 \\ \frac{1}{7} & 1 & 1 & \frac{1}{5} \\ \frac{1}{9} & 1 & 1 & \frac{1}{4} \\ \frac{1}{3} & 5 & 4 & 1 \end{bmatrix}$$

From these data, we now have the probability IL for the new populations all over the world to choose their L2. Figure 2 shows the details. The result shows that Japanese is the second choice of L2, the reason is that Japan is a country with high GDP and big culture influence on website due to animation, comic and game.

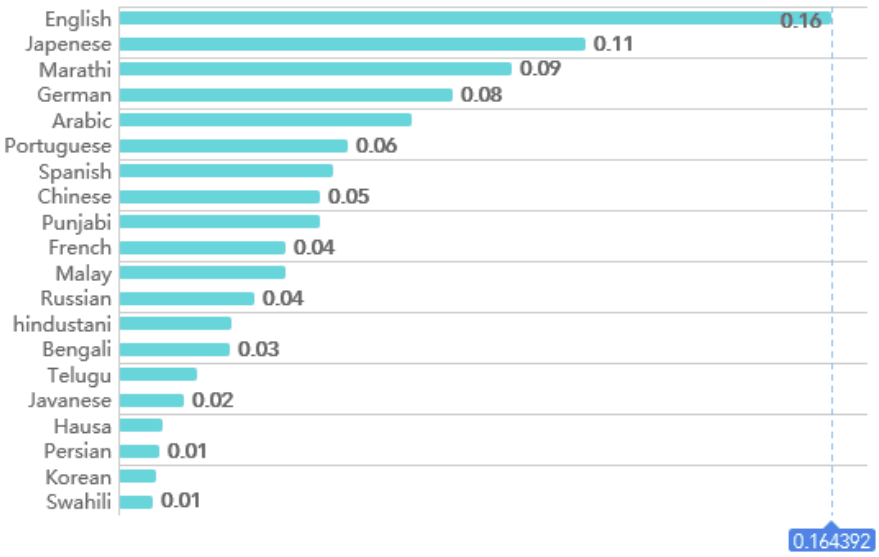


Figure 2: IL of 20 languages

Thus, S_{L2} is determined as:

$$S_{L2} = \Delta P(t) \cdot IL \tag{4}$$

where:

$\Delta P(t)$ newly-born population all over the world each year

3.3.2 Language Distribution

In this section, we discuss the geographical change of language distribution caused by immigration and emigration on the aspect of countries. A generic definition of language distribution in a country is shown by Table 2. The sum of all the numbers is equal to the population of the country.

	L1	L2	number
Native People	Official Language	Language 1	N_{n1}
		Language 2	N_{n2}
		⋮	⋮
Immigrants	Language A	Official Language	N_{IA}
	Language B		N_{IB}
	⋮		⋮

Table 2: language distribution

Immigration: Immigration means someone whose L2s are the official language now enter the country we observe. However, their L1s may be different because of their origins. For example, one whose L1 is language A now migrate to a country. The affection the one brings to the country's language distribution is reflected in parameter N_{IA}. After the one's arrival, the value of N_{IA} is increased by 1.

Emigration: Emigration means some native speakers whose L2s are different leaves the country we observe. Similar to immigration, the departure of a person whose L2 is language 1 will decrease the value of N_{n1} by 1.

Population growth: As we said before, the new populations of a country will decided their L2s by IL. Then a new population, opposite to emigration, indicates that the number of native people will increase by 1.

3.4 Test and Result

3.4.1 Test Language Speakers Model

Here we choose English as the example to test our model in predicting language speakers. Australia, Britain, Singapore, Canada, South Africa, Malaysia, U.S.A and Ireland are English-speaking countries. Add all the populations of these countries, we will obtain the number of English native speakers.

(1)Native speakers

The parameters of Britain are [7]:

P_0 6.18x10^7

K 6.9x10^7

r 0.07439

Then the perdition of British population is shown in Figure 3. Similarly, we obtain all the English-speaking countries' populations in Figure 4. The sum of them is the number of native English speakers. Repeat the processes, we can do the same to all the languages. Due to limited space, we could not draw the detail of them here. The data of the rest languages' L1 population will be shown in Figure 6 combined with L2 data below.

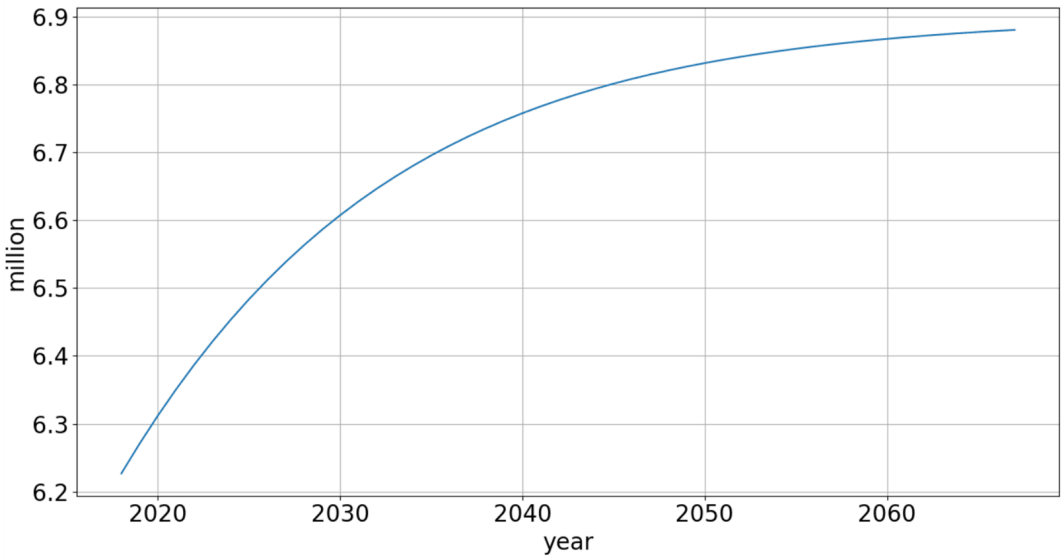


Figure 3: British population

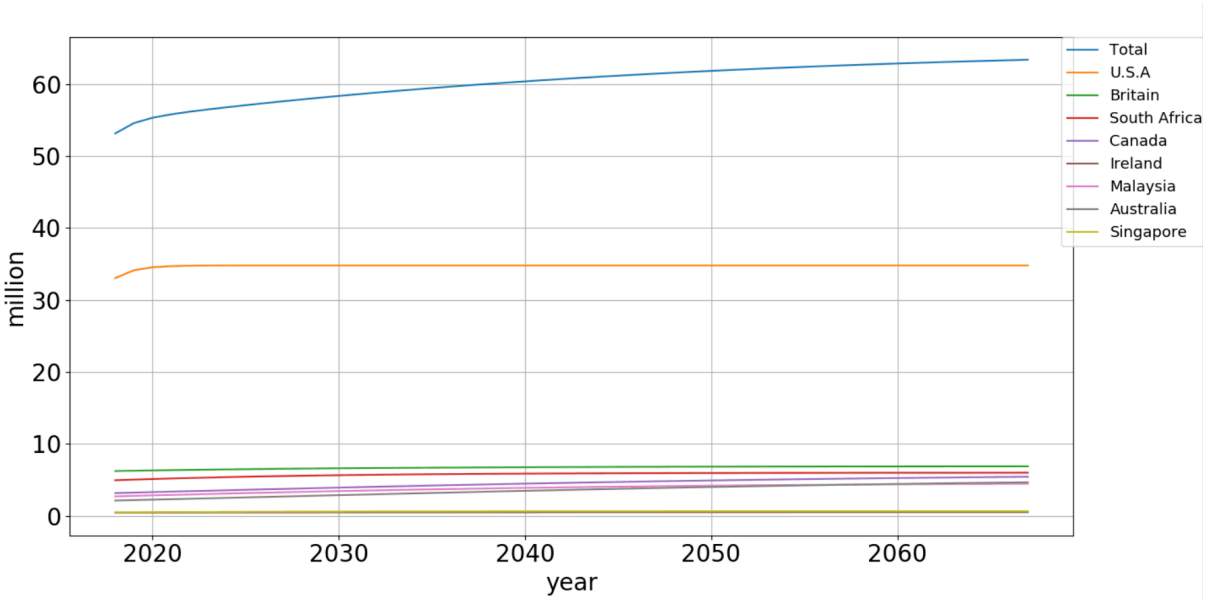


Figure 4: population of English-speaking countries

(2)Second language speakers

The number of L2 speakers of a language is divided into two parts: original speakers now exists and new ones who choose this language as L2. The data of first part can be found on the website [2]. The population of new speakers is determined by IL in Figure 2. So the data of L2 speakers is shown in Figure 5 below. From which we can find that English has a greater L2 population base and still have a strong trend to grow.

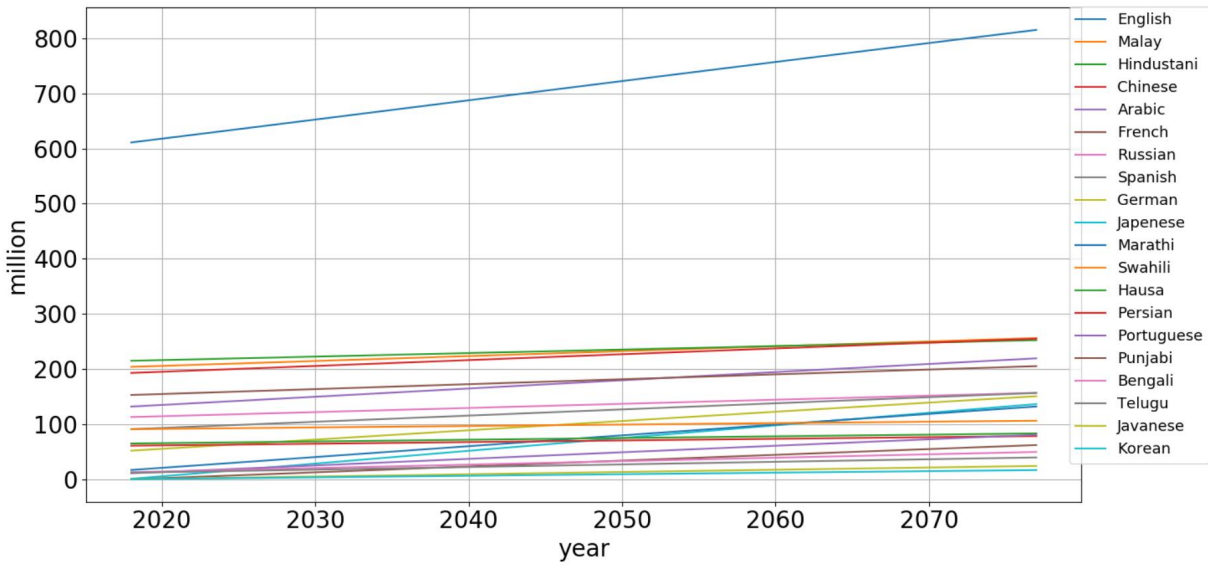


Figure 5: number of L2 speakers

(3)Total population of languages

Now that we have obtained L1 and L2 population of the languages, the prediction for total speakers of languages is now available. We draw them in Figure 6. The name list of countries right of the Figure is sorted in the order of top-twenties. Exceeding Chinese, English becomes the language with most speakers. There is no new language entering the top-ten list.

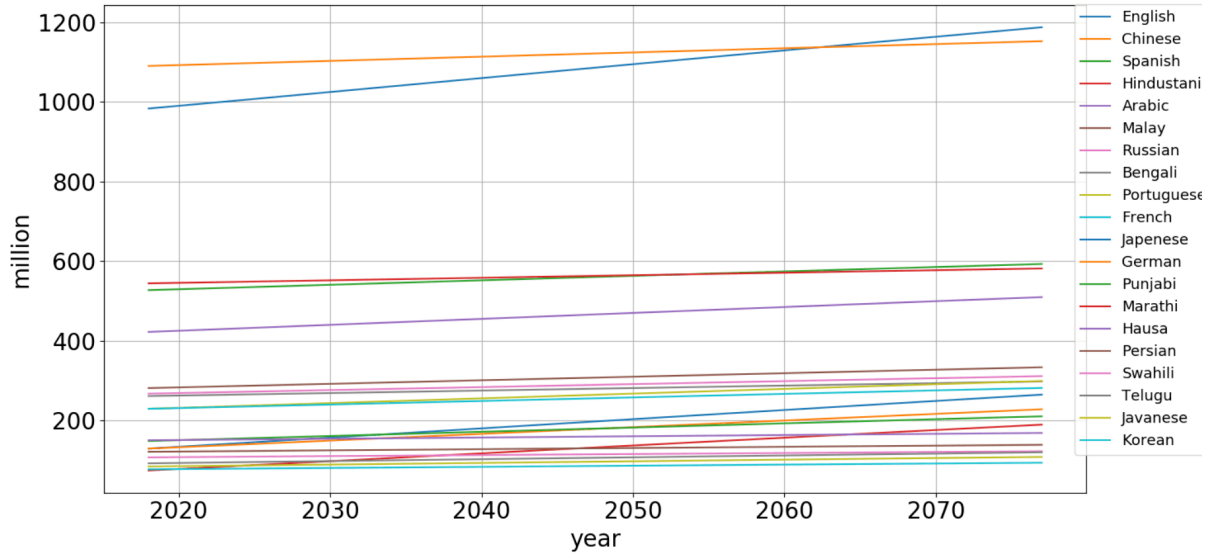
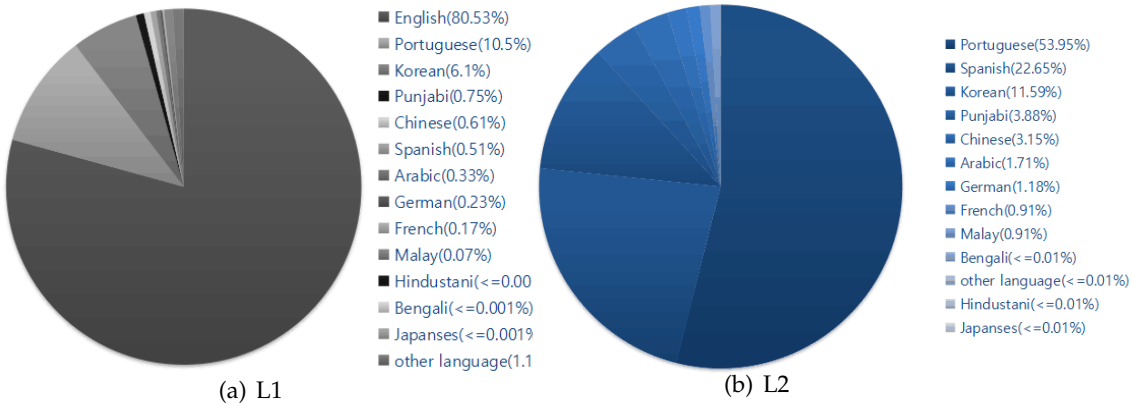


Figure 6: total population of all languages

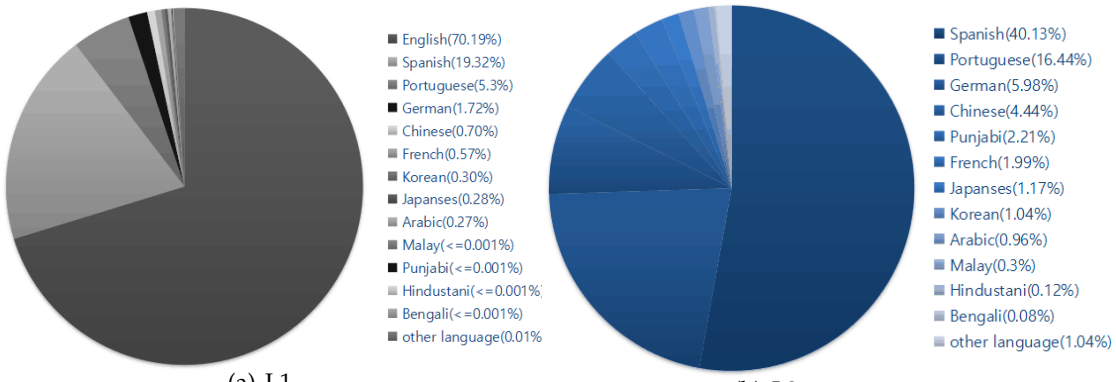
3.4.2 Test Language Distribution Model

Here we use the data of America as example to test our language distribution model. We consider immigrants from 5 countries(Mexico, Philippines, Puerto Rico, Germany and China) moving to America in the future, because the amount of immigrants from the rest countries are too small. Similarly, 5 countries(Mexico, Canada, Puerto Rico, Britain and Philippines) are considered as American emigrants' destinations. All the data is collected from United Nations [8] [9] and Wiki [2]. Today the language distribution in America is shown in Figure 7. And Figure 8 shows the distribution after 50 years according to our simulation.



(a) L1 (b) L2

Figure 7: language distribution of America



(a) L1 (b) L2

Figure 8: language distribution of America after 50 years

From the figures above, it is clear that language distribution in America have changes in both L1 and L2. On the aspect of L1, Spanish replaced Portuguese's place. Native English speakers occupy a smaller proportion because America is a country with low birth rate. Immigrants will impact the position of official language if there is no limit to immigrants in this kind of countries. What is more, Spanish also become the first choice of L2. Korean lost its top-five position on L2 list.

4 Model II - Profit Model

4.1 Overview

In the last section, we obtain the distribution of the languages used in different areas in the next 50 years. To explore the relation between the language distribution and the profit of setting up the offices, we construct a profit model to identify the most suitable sites for the offices.

4.2 Further Assumptions

- Only 30 countries with relatively high GDP are taken into consider. For the reason that multinational corporations aim to generate higher profit, we assume that building offices in countries with small economic output would violate the goal of the company.
- The proficiency of a second language would be lower than native language when a staff is newly hired. If the staff uses second language to do daily work, the proficiency of this language will gradually rise. After 2 years, the performance of second language makes no difference compared to the native one.
- An employee will be on the payroll for 10 years. New employees will be recruited in every 10 year.
- All employees in one office create equal value for the company.
- All employees use office language to communicate with each other in daily work. We consider that the "communication" in an office is not just talking to each other. All the documents, software in computer and any other office supplies should be the same in language. So they must use office language while working, ignoring other languages they might master.
- The probability of communication between any two staffs in the office is the same.
- The costs for infrastructure construction of offices are the same in any locations.
- When a new office is set up, a management team will be sent from two existing offices. Other employees will be recruited in the locality.
- Inasmuch as there exists offices in both USA and China, they will not be considered as the destination of offices when to achieve a dominant position in the world market.

4.3 Establishment

In this section we focus on determining the revenue R and cost C. Then profit P_{rf}, the net revenue, can be calculated as

$$P_{rf} = R - C \tag{5}$$

4.3.1 Efficiency

Before measuring R and C, we introduce E here to judge work efficiency in an office. As we assumed, staffs are not familiar with their second languages until they use it in daily work for two years. Then the efficiency of work between two newly hired staffs while using the same language is quantified in the following table:

Communication	E
L1 to L1	$E_1 = 1$
L1 to L2	E_2 ($E_2 < E_1$)
L2 to L2	E_3 ($E_3 < E_2$)

Table 3: efficiency of green hands

Table 3 shows that when they use the same language as their native ones to communicate, E reaches the maximum. On the contrary, when they use a language which both of them are unfamiliar with, though they can roughly understand each other, the efficiency is much lower.

Now the company already has offices in Shanghai and New York. What is more, all the staffs are required to be able to speak English. So we have two kinds of staffs divided from their native languages. When a new office is set up, we need to send a management team made up by the staffs shown in Table 4 to the new office.

Staff	L1	L2
Chinese Staff	Chinese	English
American Staff	English	uncertain language

Table 4: staffs in company now

We have assumed that the company will not set new offices in China or America temporarily. Since China is the only country in the world speaking Chinese(Singapore is classified into English-speaking countries in this paper), a new office will be set up in English-speaking countries or the other countries.

(1)Set up new office in an English-speaking country

This means that the staffs in the new office will have no difficulty in communication at the first time they meet as long as we choose American staffs to make up the management team. What ever the American staffs' L2s are, their work efficiency with local people is E_1. **So it is better to choose English as office language in English-speaking countries.**

(2)Set up new office in other countries

The structure of an office in these countries is decided by the office language. If English is chosen to be the one, there is no doubt that we choose American staffs as managers because of their proficiency of English. The structure of this case is represented in Table 5. If local language is chosen to be office language, the structure changes to Table 6.

Staff	L1	L2	number
Manager	English	any language	n_1
Local Staff	local language	English	n_2

Table 5: structure of an office using English

Staff	L1	L2	number
Manager	English	local language	n_1
Local Staff	local language	English	n_2

Table 6: structure of an office using local language

The efficiency of the two cases can be measured by

$$E_{oE} = \frac{C_{n_1}^2 E_1^2 + C_{n_1}^1 C_{n_2}^1 E_2 + C_{n_2}^2 E_3^2}{C_n^2}$$

$$E_{oL} = \frac{C_{n_1}^2 E_3^2 + C_{n_1}^1 C_{n_2}^1 E_2 + C_{n_2}^2 E_1^2}{C_n^2}$$

where:

n $n = n_1 + n_2$

E_{oE} average efficiency of an office in other countries using English

E_{oL} average efficiency of an office in other countries using local language

Because the number of managers (n_1) is always smaller than local staffs(n_2) in a company, we can conclude that E_{oL}>E_{oE} after simple calculation. **So it is better for the company to choose local language as office language in other countries.**

4.3.2 Cost

Cost of setting up an office is decided by two parameters: cost for infrastructure construction C_I and employee wages C_W. As we assumed, the first one is a constant.Here we discuss the relation between C_W and language distribution.

C_W is made up by two parts: C_{WP} - wage to one's position and C_{WL} - the wage to one's L2. C_{WP} is a constant according to one's position in the office(manager C_{WPm} or local staff C_{WPl}). C_{WL} is detrmined by the population proportion of learning a L2 we want.

For example, we need to send some American staffs to work in Japan. Their L2s should be Japanese because the office there uses Japanese to obtain highest E. Now we should pay them more because only they can do this job. Meanwhile, we should also pay the Japanese staffs for their knowledge of English. The amount of C_{WL} will be big if the population of certain L2 is small in a country. On the contrary, if we do not need the staffs to do the job requiring certain L2, we do not have to pay them C_{WL} , either.

Here we use C_{WPi} and C_{WLi} to represent the wages we give to a single staff in the office, and n is the total number of staffs. Considering the wages will rise in the future, both C_{WPi} and C_{WLi} are functions of t . Finally, the total cost of an office having set for t years can be concluded as

$$C(t) = C_I + \sum_{year=1}^{t} \sum_{i=1}^{n} (C_{WPi}(t) + C_{WLi}(t)) \tag{6}$$

4.3.3 Revenue

We introduce b to represent the value a staff brings to the company. Of course, b will change when offices are set in different countries according to the GDP per capita. What is more, b will rise with time because of inflation. Then $n \cdot b$ is the total income of an office with n staffs. However, if not all the staffs use their native language in an office, the average work efficiency E_{oi} is lower than 1 in this office in the first 2 years. Moreover, considering the tax, we introduce r_{tex} to limit the revenue every year.Then the value R of an office having set for t years brings is

$$R(t) = \sum_{year=1}^{t} E_{oi}(t) \cdot n \cdot b(t) \cdot r_{tex} \tag{7}$$

4.3.4 Profit

Based on Equation 5, 6 and 7, we can finally determine profit P_rf in Equation 8.

$$P_{rf}(t) = \sum_{year=1}^{t} \left(E_{oi}(t) \cdot n \cdot b(t) \cdot r_{tex} - \sum_{i=1}^{n} (C_{WPi}(t) + C_{WLi}(t)) \right) - C_I \tag{8}$$

4.4 Test and Result

We choose Japan here to be our example. The parameters in this simulation are:

C_{WPm}	\$67.5K per year, rise by 50% every 5 year
C_{WPl}	\$30K per year, rise by 50% every 5 year
C_{WL}	$(1 - \frac{population\ of\ certain\ L2}{total\ population}) \cdot 40000$ \$
E_2	0.8, gradually rise to 1 after two years
r_{TEX}	20%
n_1	15
n_2	85
C_I	\$20M
$b(t)$	\$80K per year, decided by Japanese GDP. Rise every 10 year

After putting data into the model, the profit in Japan results. In Figure 9, we can see the profit curve has some small hollows every 10 year. That is because we hire new staffs at that time, the new management team is made up by American staffs who can speak Japanese but not familiar with it. So the efficiency has a sudden decrease. Furthermore, we will recover our cost in 2019.

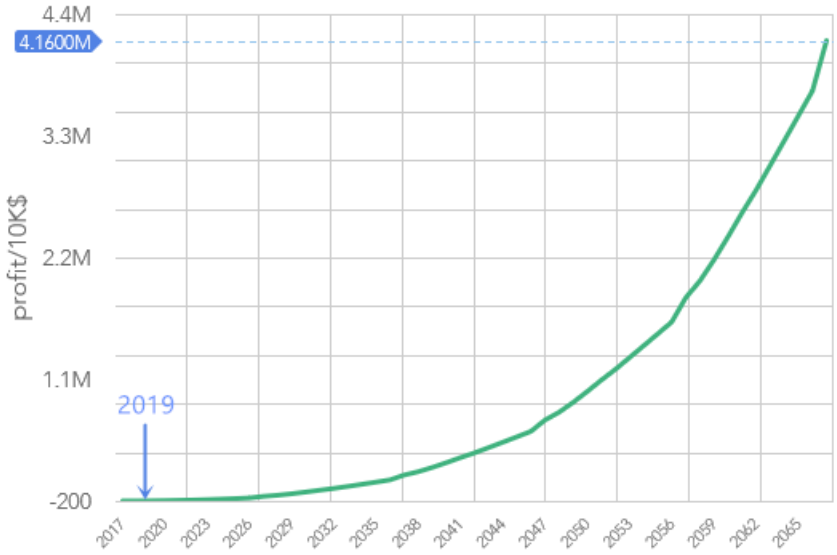


Figure 9: profit in Japan

Repeating the same process, profits of all 30 countries results. We draw the curves of top-8 countries in Figure 10. From the figure we finally decide the 6 locations with most profit 50 years after: Japan, German, France, UK, India and Russia. However, compared to other five countries, profit in Russia is obviously smaller. So we can remove Russia from the location list and build 5 new offices only.

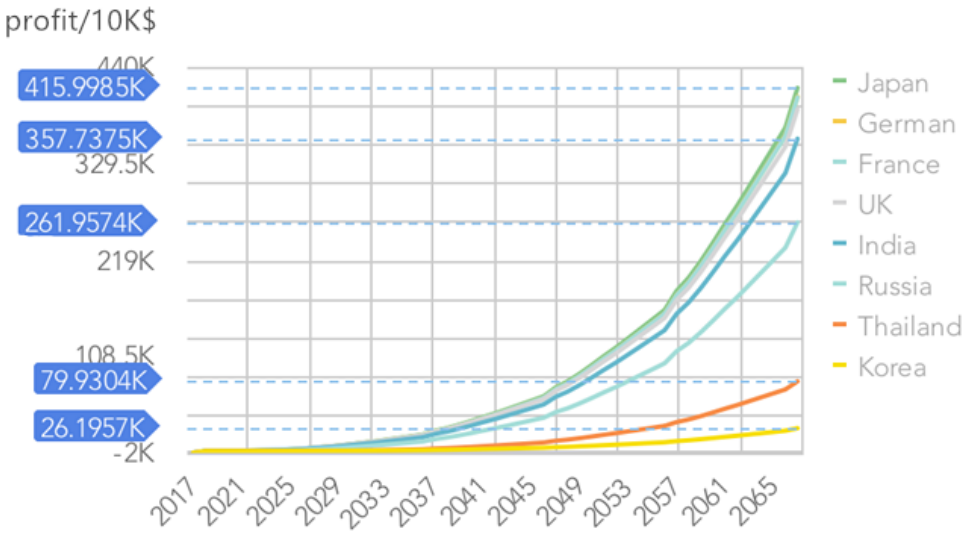


Figure 10: prediction of profits

5 Sensitivity Analysis

The number of staffs in an office: n

We set the value of n 100 in our simulation above. However, the number may change due to business strategy. Here we simulate the profit in Japan with different ns. The profit curves are shown in Figure 11. We can conclude that the influence from the change of n is small to our model. The profit becomes slightly higher when n is bigger. That is because staffs can make larger profit than their wages.

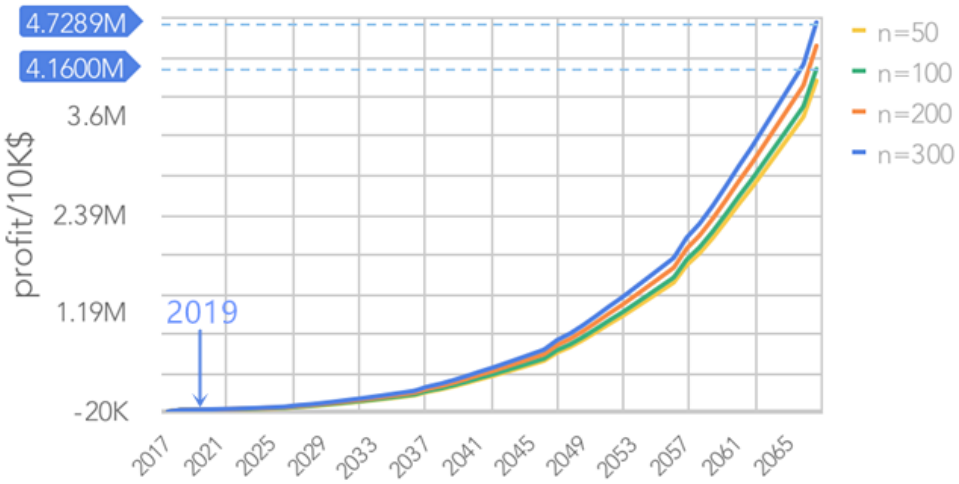


Figure 11: profit in Japan

6 Conclusions

- **The top-ten language list will slightly change after 50 years.** English will be the language with most speakers, and no new languages entering the top-ten list.
- **Immigrants and emigrants slightly change the distribution of a country.** The proportion of native speakers will gradually decrease because of immigrants in countries with low birth rate.
- **Office language should be the local language.** After our deduction and calculation of work efficiency, we determine that local language is the best choice in all the cases.

- **Japan, German, France, UK, India and Russia are the 6 locations to set offices.** These 6 countries provide most profits in all 30 countries we consider.

7 Strengths and weaknesses

Strengths:

- We took the transmission of languages on website into consider, which accords with the trend of informatization in the future.
- We quantified influence of lanauges by AHP to determine the newly-born populations' choices of second language.
- Our model is generic. We can use the models to simulate various situations with credible data.

Weaknesses:

- We did not consider national policy.
- Profit model may be inaccurate because the cost and revenue are related to some complex factors in real life.

References

[1] https://en.wikipedia.org/wiki/English_language

[2] https://en.wikipedia.org/wiki/List_of_languages_by_total_number_of_speakers

[3] https://en.wikipedia.org/wiki/Logistic_function#In_ecology:_modeling_population_growth

[4] FENG Shouping. On Prediction Model of China Population Growth.September 15,2008.http://jpkc.huanghuai.edu.cn/include/htmleditor/uploadfile/20140504040311239.pdf

[5] http://www.dii.unisi.it/~mocenni/Note_AHP.pdf

[6] Hong Zhiguo. Li Yan. Fan Zhihua. Wang Yong. Calculation on High-ranked R I of Analytic Hierarchy Process. http://www.cqvip.com/read/read.aspx?id=6341137

[7] https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/timeseries/ukpop/pop

[8] http://migrationsmap.net/#/USA/departures

[9] http://www.un.org/en/development/desa/population/migration/data/estimates2/estimates17.shtml

Memo

To: Chief Operating Officer
From: Team # 82277
Date: February 12th, 2018
Subject: Choose locations and languages of new offices.

Our memo aim to help you determine the locations for setting the offices and deciding the most efficient language to be used. As is known to all, a company sets up its new offices in order to open up international markets and get higher profit. Therefore, we only consider the countries ranked top 30 by GDP.

Suggestions for office language

We established a communication efficiency model. The higher the communication efficiency in the offices, the higher profit the company will generate. After calculating the communication efficiency of different office languages using our model, we draw the conclusion that the offices in English speaking countries should choose English as the office language. Otherwise, local language should be chosen.

Suggestion for the locations of International offices

We established a population growth model and a migration model to predict the language distribution in these countries. The salary of employees is decided by the second languages they master and the language distribution in their countries. That is to say, the fewer people speak the language as second language, the more the staffs should be paid, resulting higher expenditure of the company. It is because the profit from offices is in proportion to the communication efficiency.

We used computer to simulate our economic model in 30 countries. It turns out that in the long run, setting up international offices in Japan, Germany, France, Britain, India and Russia will help generate most profit.

Because the world economy is affected by geopolitics, pandemics and natural disasters, it is very difficult to accurately predict the income in next 50 years. Therefore, in order to avoid risk, we recommend to maximize the profit in short term. If there is a need to reduce the number of office, we recommend the removal of Russia from the list. Because Russian provides smaller profits compared with the other five countries in our simulation results.

Finally we attach a heat map here to give you a visible distribution of profits all over the world: