For office use only	Team Control Number	For office use only
T1 ________________	**72995**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**B**	F4 ________________

2018
MCM/ICM
Summary Sheet

Modeling Trends of Global Languages and Location Options for New Offices

Summary

With the globalization of economy, language plays an increasingly important role in international business communications. Based on this truth, trends of global language deserve consideration in the process of location selection for a multinational service company. However, it is not an easy thing to do.

In this paper, we propose **Speaker Prediction Model** and **Location Selection Model** to cope with this situation. By virtue of these two models, we give detailed analysis of location options according to the specific demand of a multinational service company.

Speaker Prediction Model uses study time of a language in a country to predict the distribution and numbers of various language speakers over time in different countries, based on 10,000-hour Rule. The impacts of communications between people, migration, population change and education level are considered.

Location Selection Model is an evaluation system to provide location options. There are four indices in this model, **Language Diversity Index**, **Average Costs Index**, **Market Index**, and **Distance Index**. Language Diversity Index is calculated from the output of Speaker Prediction Model, while others from data we collect. By virtue of **Analytic Hierarchy Process**, we assign different weights to the four indices out of short-term consideration and long-term consideration. Finally, **Genetic Algorithm** is implemented to optimize location options.

After the implementation of Speaker Prediction Model, we visualize distribution of various language speakers over time. A set of Bubble Maps is provided to show the impacts of population and migration. We also compare the numbers of native speakers and total language speakers in the next 50 years to predict chance of replacement.

Combination of Speaker Prediction Model and Location Selection Model results in our strategy. We mark short-term options together with long-term ones on the world map and decide the language spoken in each office respectively. Suggestions and analysis of opening fewer companies are included.

In the end, we make sensitivity analysis as well as list strengths and weaknesses of our model before conclusion.

Keywords: Analytic Hierarchy Process; Genetic Algorithm; global language trend; location option

Modeling Trends of Global Languages and Location Options for New Offices

Contents

1 Introduction

1.1 Problem Background

There are currently about 6,900 languages spoken on Earth. Despite various kinds of languages around the world, about two-thirds of the world's population share exclusively 12 native languages while other languages are claimed only by a few number of people as their native language. Figure 1 demonstrates 12 languages with the most native speakers in the world [1].

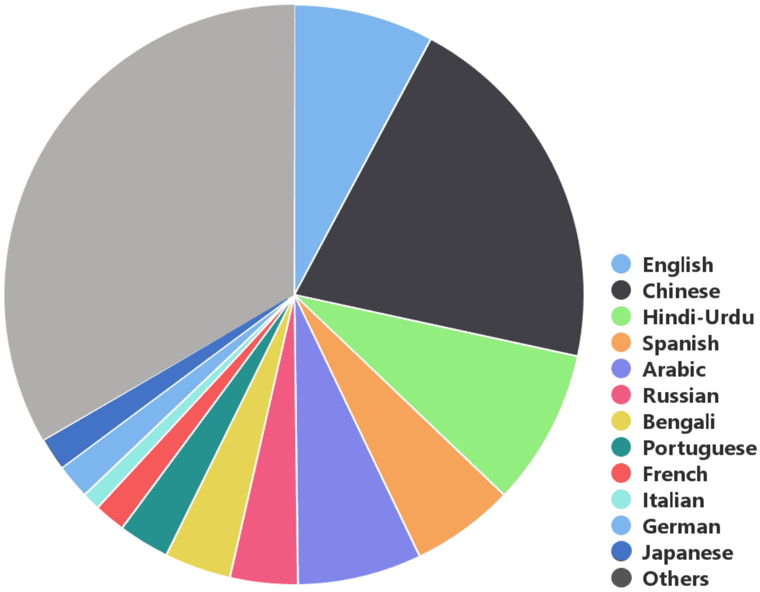

Figure 1: Languages with the most native speakers

However, apart from native languages, many people all over the world learn a second language as well. Thus, these people should be taken into consideration when it comes to comparing the total numbers of speakers of different languages. Among all the languages chosen to be a second language, English is by far the world's most commonly studied language beyond doubt. Other languages are also popular among foreign speakers to different extent.

The total number of speakers of a language may fluctuate over time influenced by a variety of factors. Additionally, nowdays in our more globalized and interconnected world, there are additional factors that allow languages that are geographically distant to interact. Many languages are spoken in more than a single country. Figure 2 shows how many countries some widespread languages are spoken in around the world [1]. The larger the word of language is, the more countries this language is spoken in.

In conclusion, the distribution and total numbers of speakers of different languages may be complicated, but to some degree predictable. Global languages play a significant role in the decisions of an international company if it plans to open oversea offices and hire employees who speak different languages.

Figure 2: Some popular languages are spoken in many countries

1.2 Our Work

In this paper we propose **Speaker Prediction Model** to predict the distribution and numbers of various language speakers over time. Taking trends of global languages into consideration, we further propose **Location Selection Model** to quantify location options. By virtue of these two models, detailed analysis is provided according to the specific demand of a multinational service company.

In Section 2, we state several basic assumptions. Section 3 contains the nomenclature used in model statement. Section 4 provides sufficient details about the two models we develop. Section 5 carrys out analysis about trends of global languages. Section 6 provides detailed strategies determining location options for new offices. At last, we further study and analyse our models in Section 7 and make some conclusions in Section 8.

2 Assumptions

Our model makes the following assumptions:

1. People in the same country are regarded as an integral whole.
2. Native language and second languages learning in the same country should be analysed in different methods.
3. In the next 50 years, it is only possible for the languages in the current top-fifteen lists to be ranked into top-ten lists because the numbers of people speaking them will not change dramatically. Thus, we will exclusively consider the countries with people speaking these 15 languages.
4. The time that people from a country spend on learning a second language on themselves is determined by the education level of the country.
5. The time that people from a country spend on communicating with others using a second language is determined by the proportion of people possessing this language in both this country and foreign countries.

6. Everyone is supposed to possess a language after he/she is born, which is his/her native language.

3 Nomenclature

In this paper we use the nomenclature in Table 1 to descibe our model. Other symbols that are used only once will be described later.

Table 1: Nomenclature

Symbol	Definition
i	The *i*th country considered in our models
j	The *j*th language in the current top-fifteen list by total speakers
n	The number of total countries considered in our models
N_i	Number of people in the *i*th country
$N_{i,j}$	Number of people possessing the *j*th language in the *i*th country
α_i	Education level of the *i*th country
β_j	Proportion of people possessing language *j* in all countries considered
$T_{i,s}$	Hours on self-learning per person in country *i*
$T_{j,c}$	Hours on communicative learning of language *j* per person in a country
$T_{i,j}$	Hours on learning of language *j* per person in country *i*
$N_{i,b}$	Number of people who are new born in the *i*th country
$N_{i,d}$	Number of people who die in the *i*th country
$N_{i,mi}$	Number of people who immigrate into the *i*th country
$N_{i,mo}$	Number of people who emigrate from the *i*th country
LD_i	Language diversity index of the *i*th country
G_i	Real GDP per capita of the *i*th country
C_i	Average costs index of the *i*th country
E_i	Engel's coefficient of the *i*th country
ρ_i	Density of population of the *i*th country
M_i	Market index of the *i*th country
D_i	Distance index of six new offices

4 Statement of Our Models

In this section, we will discuss all details about our models. These models take several fields into consideration, ranging from education level to global migration and communications. Figure 3 provides an overview of the models we apply to solving the problems. To begin with, we first develop two submodels of our **Speaker Prediction Model** to respectively capture and predict the distribution of native and second language speakers over time. Then, we take the global population and human migration patterns into consideration to improve our models. Finally, the **Location Selection Model** based on GA with weights determined by AHP is introduced to locate the new international offices of our client company considering the changing nature of global communications.

4.1 Speaker Prediction Model

Since we will possess our native language after we are born, different submodels should be applied to capture and predict the distribution of native and second language speakers over time.

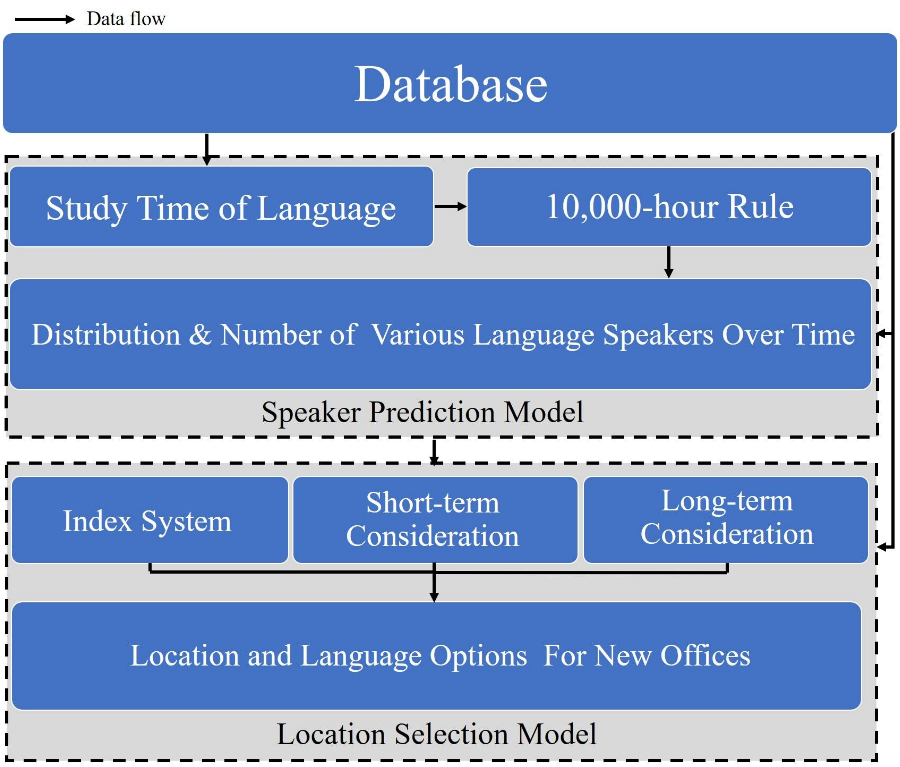

Figure 3: Overview of our models

4.1.1 Prediction of Second Language Speakers

The number of people speaking a second language in a country is determined by both domestic and foreign factors, such as the language(s) used or promoted by the government in the country, the language(s) used in schools, social pressures, and communication and assimilation of cultural groups. However, to simplify, all these factors can be reflected by just one parameter, the time people from the country spend on learning and practising the second language since it can only be spoken after acquired learning.

We define $T_{i,j}$ as the average time spent on learning a specific second language j per person in the ith country in a single day. Learning is only classified into self-learning and communicative learning in our model. Thus, we also define that $T_{i,s}$ stands for average time spent on self-learning per person in the ith country in one day while $T_{j,c}$ for average time spent on communicative learning of the jth language in a country in one day. We get the formula:

$$T_{i,j} = T_{i,s} + T_{j,c} \quad , \tag{1}$$

We assume that $T_{i,s}$ is determined by the education level of the ith country while $T_{j,c}$ is determined by the proportion of people possessing the jth language in both the ith country and other foreign countries in Equation (1). Hence, formulas are given as:

$$T_{i,s} = \frac{\alpha_i}{\overline{\alpha}} t_s \theta = \frac{\alpha_i}{\frac{\sum_{k=1}^{n} \alpha_k}{n}} t_s \theta \quad , \tag{2}$$

$$T_{j,c} = \frac{\beta_j}{\overline{\beta}} t_c = \frac{\beta_j}{\frac{\sum_{k=1}^{15} \beta_k}{15}} t_c \quad , \tag{3}$$

where

$$\theta = e^{\frac{2 - t_s}{4}} \quad ,$$

$$\beta_j = \frac{\sum_{k=1}^{n} N_{k,j}}{\sum_{k=1}^{n} N_k} \quad ,$$

and t_s is the average time spent on self-learning per person around the world while t_c is the average time spent on communicative learning per person around the world. θ is the efficiency of self-learning.

With Equation (1)–(3), we can derive the average time spent on learning a specific second language *j* per person in the *i*th country in a single day. Then we use $N_{i,j}$ to approximate the people who are learning the second language *j* in the *i*th country, so the formula that we use to determine the change of the number of people speaking the second language *j* in the *i*th country in a certain time interval t_{step} is:

$$\Delta N_{i,j} = \frac{T_{i,j} N_{i,j} t_{step}}{10000} \quad , \tag{4}$$

where the constant 10000 is the minimum hours of 'intense training' required to achieve expertise, such as a second language [2].

Add $\Delta N_{i,j}$ of all the countries considered together, we can get the change of the number of the *j*th language speakers in the time period t_{step}. Based on it, the distribution of second language speakers over time can be captured.

4.1.2 Prediction of Native Language Speakers

Every country regards one or several languages as its official language(s), which can be seen as its native language(s). The change of native language speakers in a country is more closely connected with domestic factors than foreign ones. Thus, we ignore the factor of communications with people possessing the language in foreign countries and pay more attention to the factors like the language(s) used or promoted by the government in the country and the language(s) used in schools.

As a result, if the *j*th language is the native language of the *i*th country, Equation (1)–(3) can be simplified into the formula:

$$T_{i,j} = T_{i,s} + T_{j,c} = \frac{\alpha_i}{\frac{\sum_{k=1}^{n} \alpha_k}{n}} (t_s \theta + t_c) \quad ,$$

Then, another assumption is made that people who do not possess the native language in the country are all learning it. Hence, Equation (4) is changed into the formula:

$$\Delta N_{i,j} = \frac{T_{i,j}(N_i - N_{i,j})t_{step}}{10000} \quad , \tag{5}$$

Add $\Delta N_{i,j}$ of the *i*th country which regards the *j*th language as its native language to those of countries regarding it as its second language so that the distribution of various language speakers over time can be captured completely.

4.1.3 Influence of Global Population and Human Migration Patterns

In the aforementioned **Speaker Prediction Model**, we capture and predict the distribution of various language speakers over time in a simple situation. In this section, we will take more conditions into consideration, including the birth and death rates of these countries and migration patterns [3] between these countries.

Firstly, we include the influence of global population and human migration patterns into the analysis of the distribution of native language speakers in a country. We ignore the native speakers who move into the country. Thus, Equation (5) is changed into:

$$\Delta N_{i,j} = \frac{T_{i,j}(N_i - N_{i,j})t_{step}}{10000} + \frac{T_{i,j}N_{i,b}t_{step}}{10000} - N_{i,d}\frac{N_{i,j}}{N_i} - N_{i,mo}\frac{N_{i,j}}{N_i} \quad , \tag{6}$$

The distribution of second language speakers in a country is also affected by the global population and human migration patterns. After the same analysis, Equation (4) is now given as:

$$\Delta N_{i,j} = \frac{T_{i,j}N_{i,j}t_{step}}{10000} + \frac{T_{i,j}N_{i,b}t_{step}}{10000} + N_{i,mi}\frac{\sum_{k=1}^{n} N_{k,j} - N_{i,j}}{\sum_{k=1}^{n} N_k - N_i} - N_{i,d}\frac{N_{i,j}}{N_i} - N_{i,mo}\frac{N_{i,j}}{N_i} \quad , \tag{7}$$

Considering the influence of global population and human migration patterns, our models are improved for they includes more conditions. In the following sections, these models will be applied with data to predict the distribution of various language speakers over time around the world.

4.2 Location Selection Model

In this section, we first develop our **Location Selection Model** to locate new international offices of our client company based on **Genetic Algorithm**. Then we use **Analytic Hierarchy Process** to determine the coefficients of our model, given different importance attached to different factors.

4.2.1 Factors which Influence Locations

To begin with, we select four main factors which influence the choice of locations most—diversity of languages, costs, market and distance between locations.

Firstly, we use **Language Diversity Index** to measure the diversity of languages in a country. We define it as the probability that two people selected from the population in a country at random will speak different languages. The formula is given as:

$$LD_i = 1 - \sum_{j=1}^{15} (\frac{N_{i,j}}{N_i})^2 \quad , \tag{8}$$

Because the client company desires to have the employees of each office speak both in English and one or more additional languages, we should consider the influence of English on the Language Diversity Index. Thus, Equation (8) is modified into:

$$LD_i = [1 - \sum_{j=1}^{15} (\frac{N_{i,j}}{N_i})^2][\frac{1}{2}\frac{N_{i,E} - (N_{i,E})_{min}}{(N_{i,E})_{max} - (N_{i,E})_{min}} + \frac{1}{2}] \quad , \tag{9}$$

where $N_{i,E}$ refers to the number of people speaking English in the ith country. $(N_{i,E})_{max}$ and $(N_{i,E})_{min}$ respectively stands for the maximum and minimum of it.

Secondly, costs of opening a new office in a country can be reflected by **Average Costs Index**, which is closely connected with the real GDP per capita of the country as we assume. Hence, we give the formula:

$$C_i = \frac{G_i - (G_i)_{min}}{(G_i)_{max} - (G_i)_{min}} \quad , \tag{10}$$

where $(G_i)_{max}$ and $(G_i)_{min}$ respectively stands for the maximum and minimum of G_i.

Besides, market palys an important role in the choice of locations, which can be reflected by **Market Index**. Generally, the market of a country is influenced by its density of population and people's living standard. We use Engel's coefficient to represent the people's living standard in a country, so we get the formula:

$$M_i = (1 - E_i)\frac{\rho_i - (\rho_i)_{min}}{(\rho_i)_{max} - (\rho_i)_{min}} \quad , \tag{11}$$

where $(\rho_i)_{max}$ and $(\rho_i)_{min}$ respectively stands for the maximum and minimum of ρ_i.

What's more, it is better for these six new offices to scatter all over the world for it designs to become truly international. The minimum distance between each pair of two locations of these six locations can reflect how scattered these locations are all over the world to some extent. We call it **Distance Index**. Considering current two offices in New York City and Shanghai, formula is given as:

$$D_i = \frac{d_{min}}{d_{ns}} \quad , \tag{12}$$

where d_{min} refers to the minimum distance between each pair of two locations of these six locations and d_{ns} for the distance between New York City and Shanghai.

Based on all the work we do above, then we can define an evaluation function $f(i)$ to evaluate the conditions of a country to open an office. According to Equation (9)–(12), the formula is derived:

$$f(i) = \gamma_1 LD_i - \gamma_2 C_i + \gamma_3 M_i + \gamma_4 D_i \quad , \tag{13}$$

where γ_1–γ_4 are the weights we respectively assign to the four factors to reveal the different importance of them.

Finally, we construct total evaluation function to sum up the $f(i)$ of all six countries where the new offices locate. Then we will use **Genetic Algorithm** to decide which six countries are the best locations of new offices of our client company in Section 6.1.

4.2.2 Determining Weights of Factors with AHP

The **Analytic Hierarchy Process** (AHP) is a structured technique for organizing and analyzing complex decisions, based on mathematics and psychology. Our goal is to determine the weights (γ_1–γ_4) of four previous factors with AHP. Specially considering the changing nature of global communications and our client company resources, we decide the weights as follows:

Table 2: AHP production

Factor	Weight
Language Diversity Index (γ_1)	0.0553
Average Costs Index (γ_2)	0.5650
Market Index (γ_3)	0.2622
Distance Index (γ_4)	0.1175

Table 2 shows the result of AHP and the reciprocal matrix of AHP is in Appendix B. After calculation, we get that the consistency ratio (CR) is 0.04, which is less than the standard value 0.1, meaning that the weights we derive are reasonable. We then substitute these values of γ_1–γ_4 into Equation (13) derived in Section 4.2.1.

5 Implementation of Speaker Prediction Model

5.1 Modeling Distribution of Various Language Speakers Over Time

In this section, we will use Equation (4) and Equation (5) of our **Speaker Prediction Model** to model the distribution of various language speakers over time.

To begin with, we first select 50 countries with people speaking the languages in the current top-fifteen list of total language speakers. Then we collect data (included in Appendix A) from the Internet about all countries considered, including the numbers of people speaking different languages in these countries, total population of these countries [4] and the education indexes of these countries [5]. We assume that t_s is $1h$ while t_c is $2h$ according to some literature and set t_{step} as 365 days. Substitute all these values into Equation (4) and Equation (5), and we get the change of numbers of people speaking various languages in all these countries every year using the codes in Appendix C. Bubble maps are drawn to demonstrate the change of distribution of various language speakers over time, which are shown in Figure 4 and Figure 5.

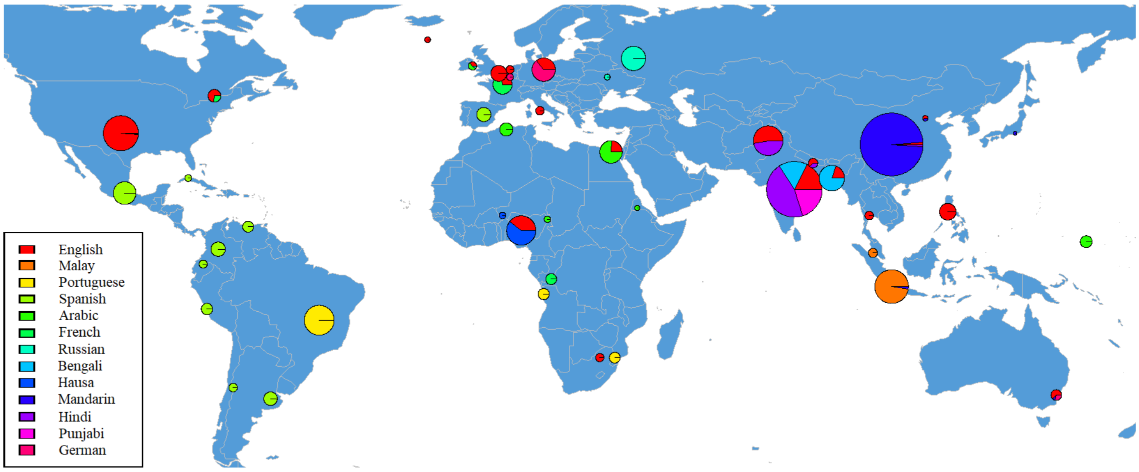

Figure 4: Distribution of various language speakers now

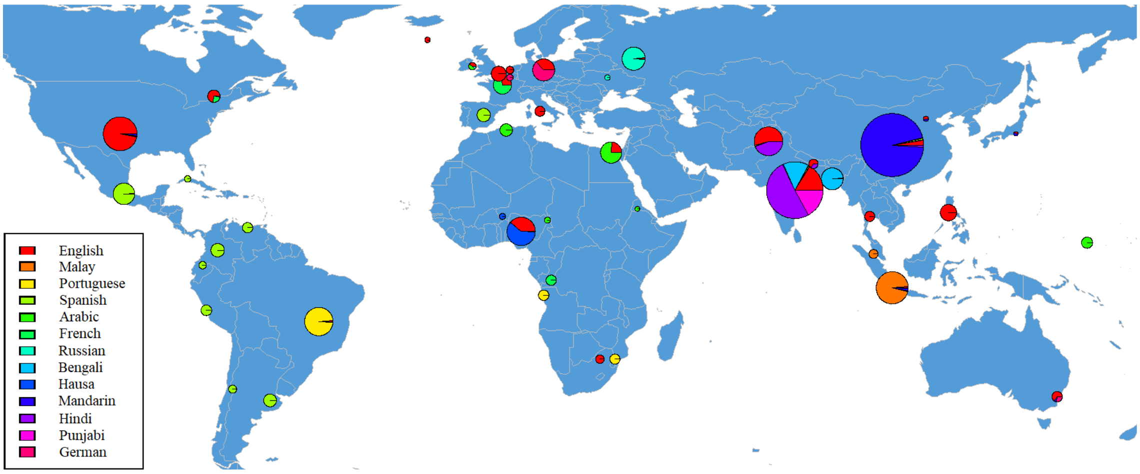

Figure 5: Distribution of various language speakers 50 years later

As we can see from Figure 4 and Figure 5, there are some changes in the languages used in several countries, such as China, India and Pakistan. More people will possess second languages in China while English will be more widespread in Pakistan. As for languages used in other countries, slight changes may happen but they will remain the same in general.

5.2 Modeling the Numbers of Native Speakers and Total Language Speakers

Add the parameter $\Delta N_{i,j}$ of all countries with people speaking the jth language together, and we can get the numbers of new total speakers of each language every year. We take the values of each language every 10 years into consideration and draw a diagram to show the trend of each language.

As is shown in Figure 6 and Figure 7, the numbers of total language speakers in the next 50 years will all increase. It seems that people are likely to become more multilingual in the next 50 years. Besides, popular languages nowadays will be learned by more people in the future in general.

We can conclude from Figure 7 that none of the languages in the current top-ten list by total speakers will be replaced by another following language in the

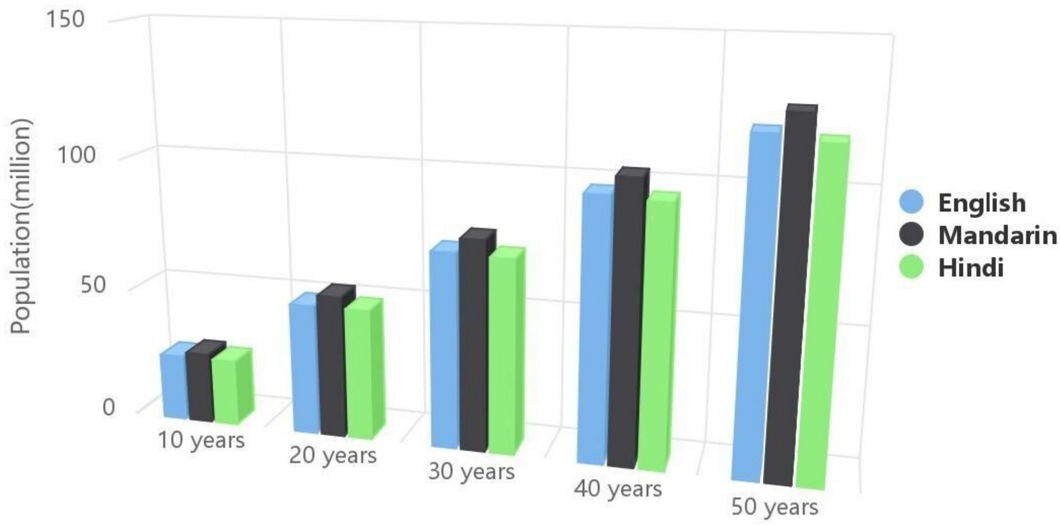

Figure 6: Numbers of people newly possessing three popular languages

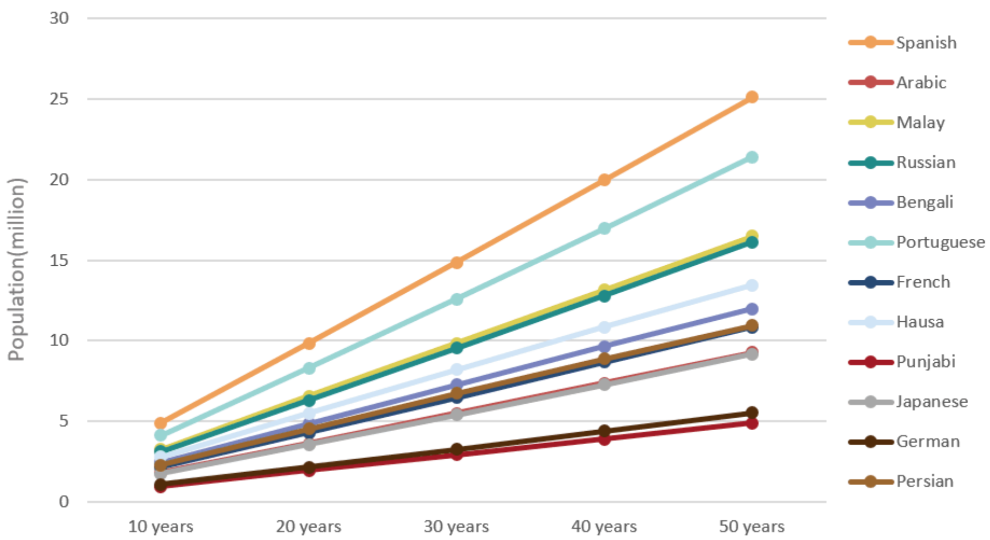

Figure 7: Numbers of people newly possessing other languages

next 50 years. We can take the 10th language French and the 11th language Hausa in the list into consideration. Even though more people will newly possess Hausa than French in the following 50 years, the difference of numbers between them is far from closing the 80 million gap between the numbers of people speaking French and Hausa at present. As a result, French will remain in the top-ten list by total speakers for a quite long time.

If we only consider the numbers of native speakers of each language, we can apply Equation (5) to the situation and draw similar diagrams to Figure 6 and Figure 7. The diagrams are as follows:

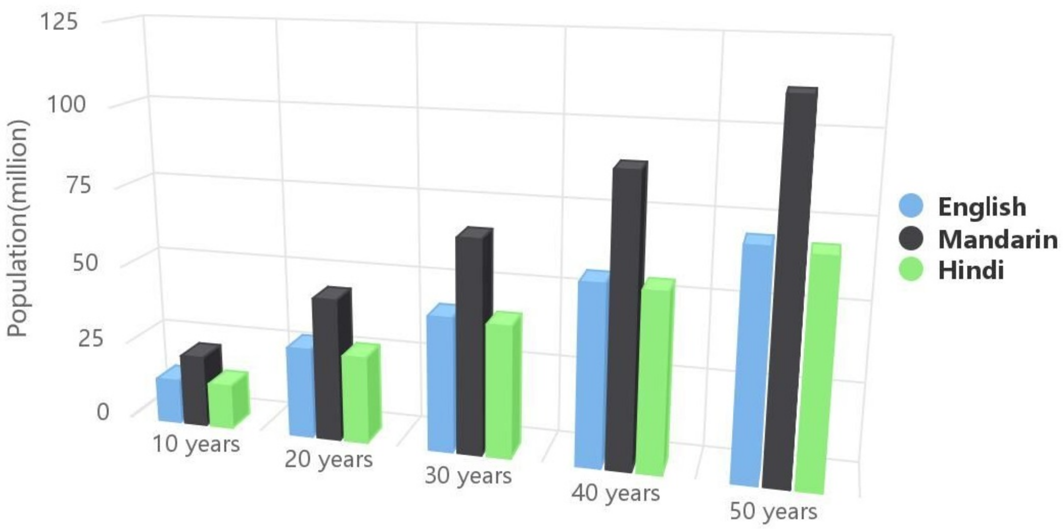

Figure 8: Numbers of native people newly possessing three popular languages

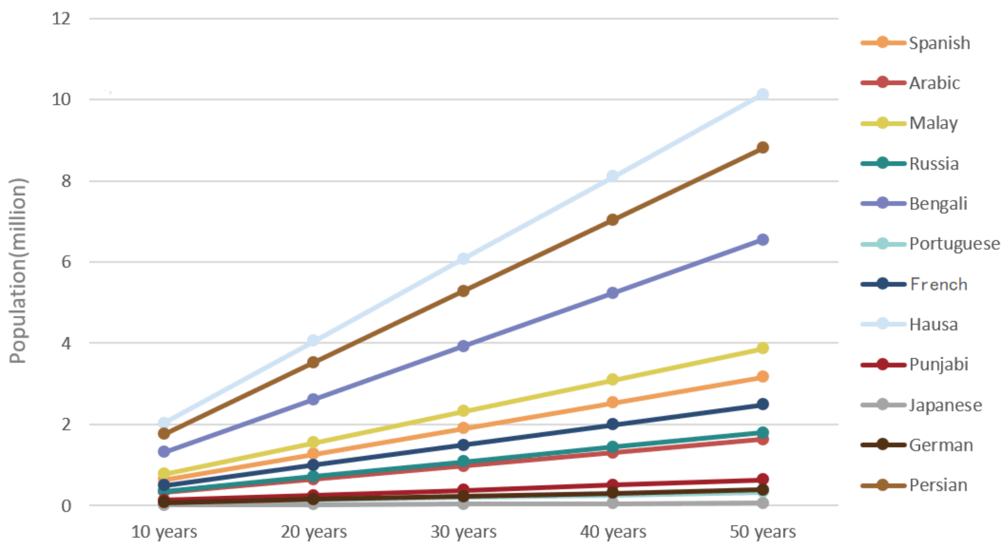

Figure 9: Numbers of native people newly possessing other languages

As is shown in Figure 8 and Figure 9, the numbers of native speakers of each language in the next 50 years will all increase. It seems that people living in a native country are more likely to speak its native language in the next 50 years.

Besides, popular languages nowadays will be learned by more native people in the future in general.

We can conclude from Figure 9 that none of the languages in the current top-ten list by native speakers will be replaced by another following native language in the next 50 years. We can take the 10th language Japanese and the 11th language Hausa in the list into consideration. Even though more native people will newly possess Hausa than Japanese in the following 50 years, the difference of numbers between them is far from closing the 43 million gap between the numbers of native people speaking Japanese and Hausa at present. As a result, Japanese will remain in the top-ten list by native speakers in the next 50 years. However, there are nearly no new native population learning and possessing Japanese in the following 50 years. Hence, it suggests that Hausa might replace Japanese in the top-ten list by native speakers in the future beyond 50 years.

5.3 Analysis of Global Population and Human Migration Patterns

To analyse the global population and human migration patterns, we search through the Internet and literature to get data of the birth and death rate (included in Appendix A) of each country and the immigration and emigration population of each country per year [6]. We then substitute these data into Equation (6) and Equation (7) and draw a similar bubble map to those drawn in Section 5.1, demonstrating the change of the geographic distributions of these languages over time.

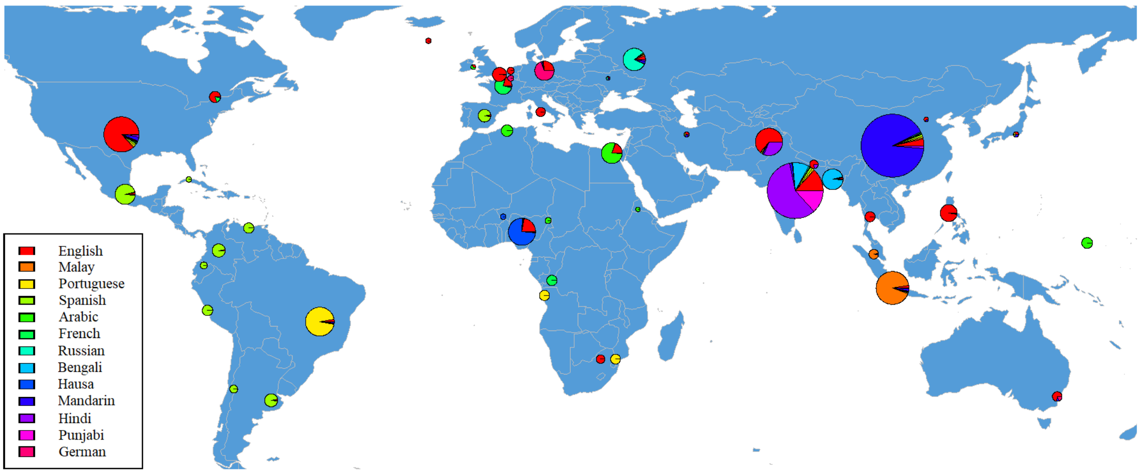

Figure 10: Distribution of various language speakers 50 years later given the global population and human migration patterns

Figure 10, compared with Figure 5, shows that the birth and death of global population and the migration between countries have a great effect on the distribution of various languages speakers. As we can see, people speaking different languages are more likely to interact with each other and live in the same country considering global population and migration. Migration makes the distribution of various language speakers more scattered all over the world.

Besides, migration also provides opportunities for popular languages to dominate and spread to more countries around the world. Hence, population speaking some unpopular languages will start to decline and they may even vanish some time in the future.

6 Strategies

In this section, we will use the **Location Selection Model** derived in Section 4.2 to develop strategies for our client company to open six new international offices. What is more, whether fewer international offices will be sufficient and reasonable as well will be discussed in this section, considering the changing nature of global communications and our client company resources.

6.1 Locating Six New Offices with GA

With the evaluation function given as Equation (13) derived in Section 4.2.1, we can use **Genetic Algorithm** (GA) to determine the best six new office locations of our client company.

To begin with, we collect data of all the countries considered through the Internet, including GDP [7], Engel's coefficient and land area [8] of each country and the distance between each two countries. Then we assume that $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \frac{1}{4}$, ignoring other factors at first. Based on the values given above, we adopt GA to determine which six countries are the most appropriate for our client company to open new international offices. In the end, we determine the located cities by choosing the most well-known cities in these chosen countries. The result of the six located cities in the short term and in the long term are both shown in Figure 11.

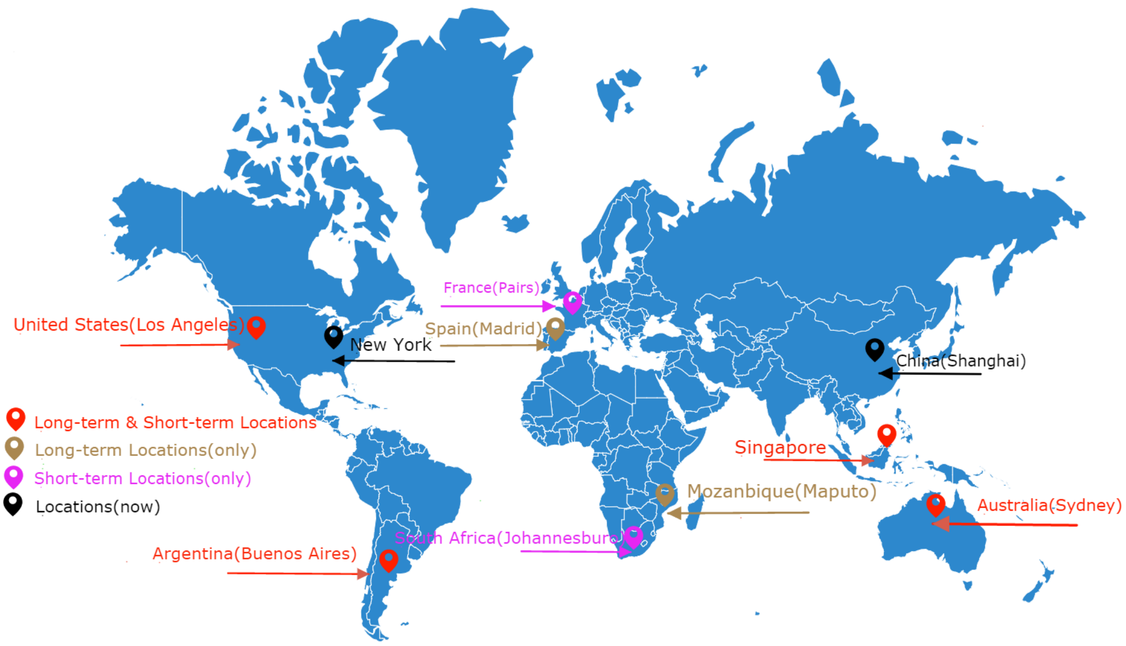

Figure 11: Six located cities in the short term versus the long term

As we can see from Figure 11, located cities in the short term versus the long term are a little bit different. However, the general principle is the same. We recommend locating at least one office in each continent and the located cities should be big cities in prime countries in the world. According to our common sense, the locations are reasonable. As for languages spoken in the offices, English is essential and another one or more languages are determined by the language(s) spoken by the most people in the located cities except English. Located cities, along with the languages spoken in the offices in the short term and in the long term are all shown in Table 3.

Table 3: Six located cities along with additional language(s) spoken in the offices except English in the short and long term

Short term		Long term	
City	Language	City	Language
Los Angeles	Mandarin	Los Angeles	Mandarin
Buenos Aires	Spanish	Buenos Aires	Spanish
Singapore	Mandarin, Malay	Singapore	Mandarin, Malay
Sydney	Mandarin	Sydney	Mandarin, Malay
Paris	French	Madrid	Spanish
Johannesburg	Hindi	Maputo	Portuguese

6.2 Discussing the Probability of Fewer Offices

Considering the changing nature of global communications and in an effort to save our client company resources, we just need to adopt the weights (γ_1–γ_4) of factors derived with AHP in Section 4.2.2 instead of collecting any additional information. These weights are derived under the condition of attaching more importance to the factor of costs and less importance to language diversity because fewer costs can save our client company resources while more frequent global communications drive people to become more multilingual all over the world, making the language diversity of countries tend to be no different.

Substitute these weights into Equation (13) and derive the maximum sum of $f(i)$ of six offices versus five offices. We find that the maximum sum of the evaluation function $f(i)$ of five offices is larger than that of six offices, of which the value is 10.9944 versus 9.708, indicating that five offices strategy is prefered than six offices strategy. We draw the result of the five located cities in Figure 12.

Figure 12, compared with Figure 11, demonstrates that only one office is located in United States in five offices strategy. We would suggest that our client company should open five new international offices rather than six ones for there are also at least one office in each continent in five offices strategy and these five offices are also scattered in the big cities of prime countries all over the world. As for the missing office in United States in Figure 12 compared with Figure 11, more frequent and convenient communications through the Internet throughout America can connect West America better with East America so there is no need opening two offices in United States. In conclusion, we strongly advise our client company to open just five new international offices rather than six offices.

The five located cities, along with languages spoken in the offices, are shown in Table 4.

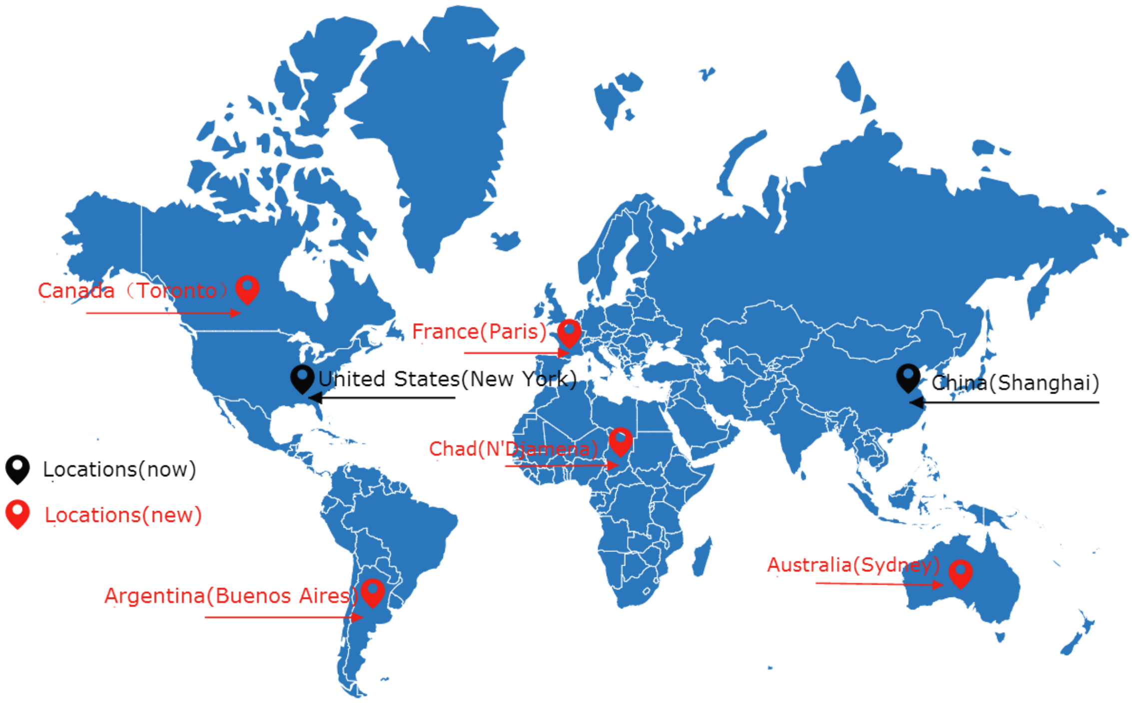

Figure 12: Five located cities

Table 4: Five located cities along with additional language(s) spoken in the offices except English

City	Language
Buenos Aires	Spanish
Toronto	Mandarin
N'Djamena	Arabic, Spanish
Sydney	Mandarin, Malay
Paris	French

7 Model Analysis

7.1 Sensitivity Analysis

Our **Speaker Prediction Model** contains several parameters. We determine some of the parameters through the search of Internet, some of them by knowledge in the literature and other methods. In this section, we would like to produce a sensitivity analysis to show whether our model is sensitive to different values of parameters.

We determine the average time spent on self-learning and communicative learning per person around the world—t_s and t_c—by ourselves. Hence, we will provide a sensitivity analysis of these two parameters in this section.

7.1.1 Impact of Average Self-learning Time t_s

t_s is set to be $1h$ in our model and not changed ever since. According to the literature, t_s is approximately this value and may change in a range. Thus, we vary this value from $0.5h$ to $3h$ and draw a diagram to demonstrate how the increased number, $\Delta N_{i,j}$ of both total language speakers and native speakers of

English varies with the change of the average self-learning time t_s.

Figure 13 shows that the larger t_s is, the more total language speakers and native speakers will start to speak English but the less influence the change of t_s will have on the increased population in the meantime. Therefore, we come to the conclusion that our model is sensitive to t_s.

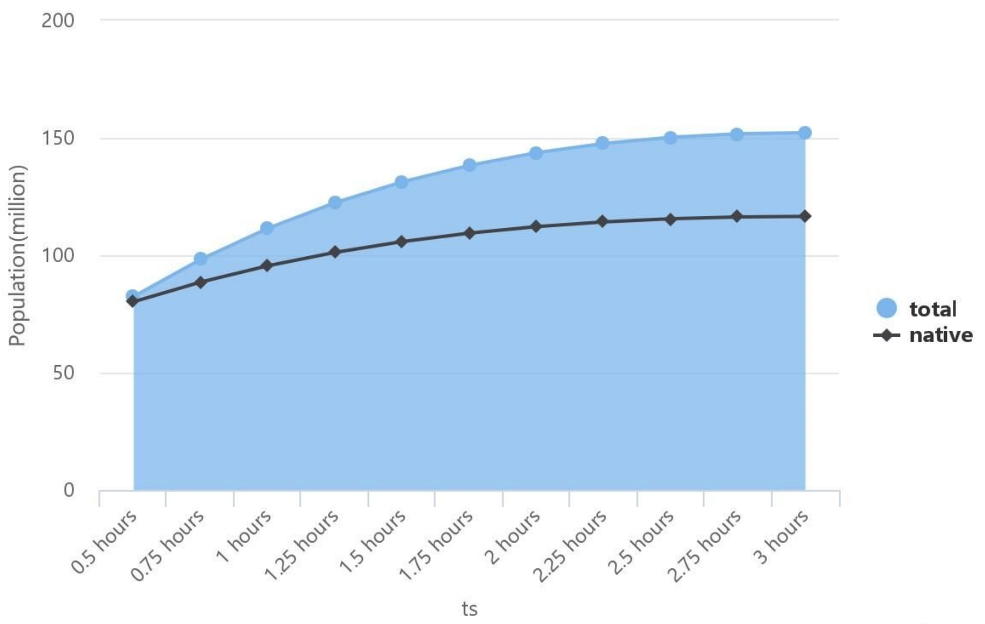

Figure 13: Impact of average self-learning time t_s

7.1.2 Impact of Average Communicative Learning Time t_c

t_c is set to be $2h$ in our model and not changed ever since. According to the literature, t_c is approximately this value and may change in a range. Thus, we vary this value from $1.5h$ to $4h$ and draw a diagram to demonstrate how the increased number, $\Delta N_{i,j}$ of both total language speakers and native speakers of English varies with the change of the average communicative learning time t_c.

Figure 14 shows that the larger t_c is, the more total language speakers and native speakers will start to speak English. The change of t_c will have the same influence on the increased population in this range. Therefore, we come to the conclusion that our model is sensitive to t_c.

7.2 Strengths and Weaknesses

7.2.1 Strengths

1. 10,000-hour rule is deeply researched by scientific community, which acts as the basis of **Speaker Prediction Model**.

2. Combination of Analytic Hierarchy Process and our index system enables **Location Selection Model** to give a comprehensive analysis according to changing demand.

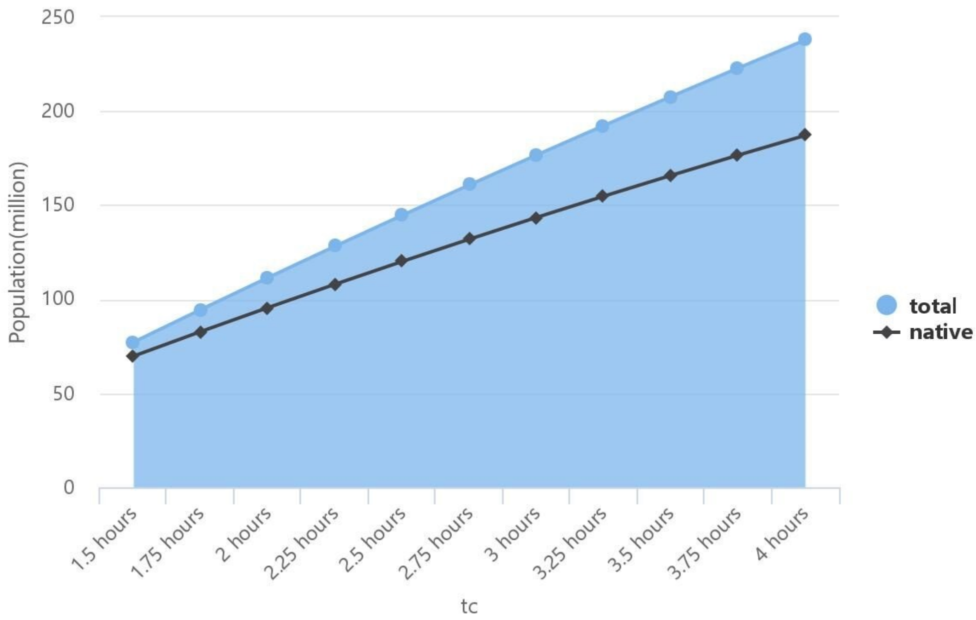

Figure 14: Impact of average communicative learning time t_c

3. We take advantage of reliable database to collect large amounts of data for the implementation of our model.
4. We draw vivid graphs such as a word cloud, bubble maps, and a flow diagram to visualize our data and results.
5. Sensitivity analysis focuses on t_s and t_c, which are two important parameters in **Speaker Prediction Model**.

7.2.2 Weaknesses

1. Due to the complexity of the problem itself, we neglect some details in our **Speaker Prediction Model**. Instead, we try to predict the distribution and numbers of various language speakers over time macroscopically, which may result in some minor deviation of our result.
2. We use Generic Algorithm in optimization process of **Location Selection Model**, which introduces randomness.

8 Conclusion

Our paper provides a detailed analysis of predicting distribution of various language speakers around the world and locating new international offices for our client company. We propose **Speaker Prediction Model** to describe the change of the distribution of various language speakers over time. We use our **Location Selection Model** to choose the most appropriate cities as our suggested locations for our client company offices. To evaluate our strategy, we define the evaluation function.

Based on real data, our models produce the distribution of various language speakers under different conditions. With the definition of evaluation function, we control our choices of new office locations to optimize our evaluation function and determine the best strategy.

Strategies are fully discussed when situations change. The impact of global communications and our client company resources are considered. Change of the number of new international offices is also included.

Memo

To: Chief Operating Officer

From: Team 72995

Subject: Trends of Global Languages and Location Options for New Offices

---

Hired to investigate trends of global languages and location options for new offices, our team respectively develop **Speaker Prediction Model** and **Location Selection Model** to solve these two parts of the problem. Then we implement our models based on real data we collect from the Internet and some literature. This memo will provide you with the results of our models and our recommendations for new international office locations.

To begin with, we first classify total speakers of one language into two different categories, native language speakers and second language speakers. Two different submodels of our **Speaker Prediction Model** are produced to separately describe and predict the distribution of them because one learns to speak his/her natvie language and second language in different ways.

In our **Speaker Prediction Model**, we define time period $T_{i,j}$ as the average time people in the *i*th country spend on learning the *j*th language. Combined with 10,000-hour rule which means that one needs at least 10,000 hours to learn a language, $T_{i,j}$ is used to determine how many people in one country will newly possess one language in a certain time period t_{step}. Based on it, we can predict the distribution and numbers of various language speakers around the world in the next 50 years considering global population and human migration patterns.

We substitute real data we collect into our models and get the following results about trends of global languages over time:

1. More languages will be spoken in most countries with more frequent global migration and communication.
2. Some currently popular languages, such as English will spread to more countries and be spoken by more people as a second language around the world.
3. The numbers of native language speakers, as well as total language speakers of most languages tend to increase in the next 50 years, indicating that people may become more multilingual in the future.
4. Even though the numbers of speakers of some languages may increase faster than some other languages, none of the languages in the current top-ten lists (either native speakers or total speakers) will be replaced by following languages given the enormous gaps between them at present.
5. Global population and human migration patterns have great influence on the geographic distributions of these languages over this same period of time, driving languages to scatter all over the world and allowing popular

languages now to become more dominant versus unpopular languages in the future.

Based on our modeling above, we develop another model called **Location Selection Model** to determine location options for new offices of your company. The evaluation system of our model includes four factors, Language Diversity Inde, Average Costs Index, Market Index, and Distance Index which we assume will have the most influence on location options. We define an evaluation function $f(i)$ with weights assigned to these four factors to evaluate how appropriate a country is for opening new offices. With countries determined, we then choose the most well-known cities in these countries to locate the new offices of your company.

If you would like to open six new international offices, we attach the same weight to the aforementioned four factors and determine the best six cities for new offices of your company by maximizing the sum of the evaluation function $f(i)$ of six located cities with Genetic Algorithm. Current data and data predicted in the next 50 years are respectively adopted for short-term and long-term consideration. With cities determined, we select the most spoken language(s) except English in the cities to be spoken together with English in the offices.Our recommendations are as follows:

Short term		Long term	
City	Language	City	Language
Los Angeles	Mandarin	Los Angeles	Mandarin
Buenos Aires	Spanish	Buenos Aires	Spanish
Singapore	Mandarin, Malay	Singapore	Mandarin, Malay
Sydney	Mandarin	Sydney	Mandarin, Malay
Paris	French	Madrid	Spanish
Johannesburg	Hindi	Maputo	Portuguese

Considering the changing nature of global communications and in an effort to save your resources, we attach more weight to the factor of costs and less to that of language diversity with Analytic Hierarchy Process because more frequent global communications drive people to become more multilingual, making the language diversity of cities tend to be no different. Then we discuss the possibility of fewer new international offices with GA in the same method above. Based on the results derived, we recommend that you should open just five new international offices as follows:

City	Language
Buenos Aires	Spanish
Toronto	Mandarin
N'Djamena	Arabic, Spanish
Sydney	Mandarin, Malay
Paris	French

Our recommends are reasonable because we locate at least one city in each continent together with two of your current offices in New York City and Shanghai and the located cities are all the big cities in prime countries in the world. We believe locating these new international offices around the world is bound to make your company become truly international and successful.

References

[1] Rick Noack and Lazaro Gamio. The world's languages, in 7 maps and charts. https://www.washingtonpost.com/news/worldviews/wp/2015/04/23/the-worlds-languages-in-7-maps-and-charts/?utm_term=.9f27ed65369a, 2015.

[2] Sarah Elaine Eaton. How long does it take to learn a second language?: Applying the 10,000-hour rule as a model for fluency. Technical report, Language Research Centre, University of Cagary, 4th Floor, Craigie Hall 2500 University Dr. NW Calgary, Alberta, T2N 1N4 Canada, February 2011.

[3] W. Neil Adger, Nigel W. Arnell, Richard Black, Stefan Dercon, Andrew Geddes, and David S. G. Thomas. Focus on environmental risks and migration: causes and consequences. *Environmental Research Letters*, 10:1–6, 2015.

[4] Population mid-2017 (millions). http://www.worldpopdata.org/map, 2017.

[5] Education index. https://en.wikipedia.org/wiki/Education_Index, 2017.

[6] Migration numbers. http://migrationsmap.net/#/USA/arrivals, 2007.

[7] Gdp of countries. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD, 2018.

[8] Land area of countries. https://data.worldbank.org/indicator/AG.LND.TOTL.K2?view=chart, 2018.

Appendices

Appendix A Data of All Countries Considered

Country	Education Index	Move in(M)	Move out(M)	Birth(‰)	Death(‰)
United States	0.89	0.94	0.071	14	8
India	0.473	0.295	0.3345	23	7
Pakistan	0.372	0.1	0.144	30	7
Nigeria	0.425	0	0.038	39	12
Philippines	0.61	0	0.1375	26	5
United Kingdom	0.86	0.1	0.1675	13	9
Germany	0.884	0.2	0.146	8	10
Canada	0.85	0.12	0.052	11	7
France	0.816	0.2	0.059	13	9
Australia	0.927	0.11	0	14	7
Italy	0.79	0.03	0.1275	10	10
Egypt	0.573	0	0.08	25	6
Thailand	0.608	0	0.02	15	9
Netherlands	0.894	0.05	0.02	11	10
Nepal	0.452	0	0.04	23	7
South Africa	0.695	0.035	0.02	21	10
Sweden	0.83	0.0325	0	10	9
Israel	0.854	0.063	0.03	10	14
New Zealand	0.917	0.025	0.02	21	15
China	0.61	0	0.2275	12	7
Singapore	0.768	0.026	0	10	4
Indonesia	0.603	0	0.07	21	5
Mexico	0.638	0.017	0.47	20	5
Colombia	0.602	0	0.065	20	6
Spain	0.794	0.058	0.04	5	6
Argentina	0.783	0.0615	0	18	8
Peru	0.664	0	0	18	8
Venezuela	0.682	0	0	18	8
Chile	0.746	0	0	18	8
Ecuador	0.594	0	0.025	18	8
Cuba	0.743	0	0.045	18	8
Algeria	0.643	0	0.085	20	12
Chad	0.256	0	0	26	5
Eritrea	0.228	0	0	26	10
Morocoo	0.468	0	0	20	10
Malaysia	0.671	0.0755	0	21	5
Russia	0.78	0.525	0.46	6	6
Ukraine	0.796	0.313	0.258	6	6
Bangladesh	0.447	0	0.315	10	6
Brazil	0.661	0.0195	0	17	6
Angola	0.474	0	0	17	6
Mozambique	0.372	0	0.02	38	15
Congo	0.511	0	0.02	39	13
Niger	0.198	0	0.02	20	13
Punjabi	0.423	0	0	23	77
Japan	0.808	0.052	0.05	9	9
Belgium	0.812	0.037	0	10	13
Iran	0.683	0.111	0.02	20	5
Afghanistan	0.365	0	0.115	20	7

Appendix B Reciprocal Matrix of AHP

	LD_i	C_i	M_i	D_i
LD_i	1	1/7	1/5	1/3
C_i	7	1	3	5
M_i	5	1/3	1	3
D_i	3	1/5	1/3	1

Appendix C Distribution of Various Language Speakers Over Time

```
clear
clc
%open data sheet
dq2=xlsread('dq2','Sheet1');
%initialize the language_country table
l_c=zeros(50,15);
l_c2=zeros(50,15);
%initialize the table of the different between years after and now
different_language1=zeros(5,15);
different_language2=zeros(5,15);
for i=1:50
    for j=1:15
        l_c(i,j)=dq2(i,j);
        l_c2(i,j)=dq2(i,j);
    end
end
%sum each language
sum_language=sum(l_c);
%sum each country
sum_country=sum(l_c,2);
Beta_language=zeros(1,15);
%acquire brta_parameter
for i=1:15
    Beta_language(1,i)=sum_language(1,i)/sum(sum_country);
end
Beta_language;
Alpha_language=...
    [0.3,0.1,0.03,0.1,0.03,0.05,0.07,0.01,0.1,0.04,0.01,0.01,0.04,0.02,0.01];
%record the max language in each country
country_max=max(l_c,[],2);
%open the data of the population in each country
sum_country2=xlsread('renkou','Sheet1');
l_c3=l_c;
iii=1;
country_max2=max(l_c,[],2);
%predict 50 years later
for i=0:49
    for j=1:50
        %enumerate 15 languages
        for k=1:15
            %find the seconds/thirds language
            if l_c2(j,k)~=country_max(j)
                l_c(j,k)=l_c(j,k)+sum_country(j,1)*0.001*10*...
                    (Beta_language(1,k)*1+Alpha_language(1,k)*2)./10000*365;

            end

            %find native language
            if l_c2(j,k)==country_max(j)
                l_c(j,k)=l_c(j,k)+(sum_country2(j,2)-...
                    country_max2(j)+0.1)*2*0.1*(sum_country2(j,3))./10000*365;

                different_language2(iii,k)=different_language2...
                    (iii,k)+(sum_country2(j,2)-sum_country2(j,1)+0.1)*...
                    2*0.1*(sum_country2(j,3))./10000*365;

            end

        end
    end
    %update the following paramater
    country_max2=max(l_c,[],2);
    sum_language=sum(l_c);
    sum_country=sum(l_c,2);
```

```
    for p=1:15
           Beta_language(1,p)=sum_language(1,p)/sum(sum_country);
    end
    %every 10 years record data
    if mod((i+1),10)==0
        different_language((i+1)/10,:)=sum(l_c)-sum(l_c2);
        iii=iii+1;
        if iii~=6
        different_language2(iii,:)=different_language2(iii-1,:);
        end

    end
end
sum_language2=sum(l_c);
sum_language2=sum(l_c);
%print the different between now and years after
different_language
different_language2
-different_language2+different_language
l_c4=l_c
for i=1:50
    for j=1:15
        if l_c4(i,j)<=0.2
        l_c4(i,j)=0;
        end
    end
end
xlswrite('shujv.xlsx',l_c4,'Sheet1','A1')
```