

For office use only

Team Control Number

For office use only

T1 \_\_\_\_\_

**77238**

F1 \_\_\_\_\_

T2 \_\_\_\_\_

F2 \_\_\_\_\_

T3 \_\_\_\_\_

Problem Chosen

F3 \_\_\_\_\_

T4 \_\_\_\_\_

**B**

F4 \_\_\_\_\_

---

**2018**

**MCM/ICM**

**Summary Sheet**

## **Summary**

Thousands of languages are spoken around on the Earth, and over 40% of people take one of the top ten most spoken language as their mother tongue. Meanwhile, more and more people are learning a second language to meet the rapid development of economy and globalization.

Aimed to predict the distribution of widespread language speakers over the next 50 years, we build a Language Development Model, and apply it to countries with a population over 10 million to describe the possibility of the new generation in a certain country studying a certain second language. We employ the Analytical Hierarchy Process (AHP) to determine the specific parameters, and carry out stochastic simulation on Python to produce the results, which shows that mild change takes place on the list of popular languages – only two of the top-10 languages are replaced, while the total number of speakers of several languages gains an impressive increase.

Based on the language distribution model we obtain, we develop a Location Determination Model. The dual-solution model further provide a recommendation for a large multinational service company to select the location of their new international office. One of the solution is oriented by language structure, and the other by geographic condition. Taking language distribution as well as economic development of different countries into account, the model suggests that new offices should be located in Nigeria, India, Germany, Italy, Brazil, Australia for short-term interest, and Nigeria, Russia, Germany, Poland, Brazil, Australia for long-term development.

The sensitivity analysis shows the strong robustness of our model. Variation of key parameters causes moderate changes to the computing result. When we test the sensitivity, we indirectly validate our model's conformance with reality that the direction of fluctuation of the result matches situations in real world. Meanwhile, we further discuss the impact of reducing the number of new office, and provide practicable advice on location determination of new offices.

**Key Words:** Language Development Model, AHP, Fuzzy Clustering, P-Center Model

## Content

|   |    |
|---|----|
| 1. Introduction.....  | 3  |
| 2. Assumptions and Justification.....                         | 3  |
| 3. Notation.....  | 4  |
| 4. Model.....   | 4  |
| 4.1 Language Development Model.....                           | 4  |
| 4.1.1 Judgment Process.....                                   | 4  |
| 4.1.2 Stochastic Simulation Process.....                      | 7  |
| 4.2 Location Determination Model.....                         | 11 |
| 4.2.1 Solution Oriented by Language Structure.....            | 11 |
| 4.2.2 Solution Oriented by Geographic Condition.....          | 13 |
| 5. Sensitivity Analysis.....                                  | 17 |
| 5.1 Sensitivity Analysis on Language Development Model.....   | 17 |
| 5.1.1 Net Birth Rate Impact.....                              | 17 |
| 5.1.2 GDP Growth Rate Impact.....                             | 17 |
| 5.1.3 Migration Rate Impact.....                              | 18 |
| 5.2 Sensitivity Analysis on Location Determination Model..... | 19 |
| 6. Further Discussion.....                                    | 20 |
| 7. Strengths and Weaknesses.....                              | 20 |
| 7.1 Strengths.....  | 20 |
| 7.2 Weaknesses.....   | 21 |
| 7. Conclusion.....  | 21 |
| Reference.....  | 21 |
| Memo to the Service Company.....                              | 22 |

# 1. Introduction

Language is the most important communication tool for human beings. Today, over 6000 languages are spoken on Earth, which help preserve and inherit human civilization and achievements. While everyone has his/her own native language which seldom changes all over his/her life, the number of people speaking a second language keeps increasing not only because of domestic factors such as school studying or governmental promotion, but also international factors including growing migration, booming global tourism and close communication online. Considering both native and second language speakers, great changes take place in language distribution and the number of speakers.

To investigate how the commonly used languages scatter worldwide, we develop a Language Development Model to forecast the future development of the 16 most spoken languages counted in 2017. Based on the original distribution of different languages and the Net Birth Rate of these countries, we build the model to simulate how language distribution changes over the next 50 years.

After obtaining the result from the Language Development Model, we build a Location Determination Model using two different solutions. For the former one, we group the countries based on their language structure. For the latter one, they are separated according to their geographic condition. After the grouping process, we set two goals to filter an optimal location for the new office in each group, namely GDP distribution and distance from other countries in the same group. Besides, we give recommended spoken language in these new offices.

## 2. Assumptions and Justification

We make some general assumptions to simplify our model. These assumptions together with corresponding justification are listed below:

- **All people take a certain language as their native language, but not every of them speaks a second language.**

Language is an indispensable tool for every person, so all people can speak at least one language, either commonly or rarely used. However, whether one speaks an additional second language depends on multiple reasons, for example, the developmental level of his/her country.

- **All existing people won't change their native and second language.**

People usually speak the same native language through his/her life. We do not consider change in native language caused by migration or other factors in the simplified model. The influence of migration will be discussed in 4.1.1.

- **The new generation all take their parents' native language as their own native language. However, their choice of the second language varies for several reasons discussed in 4.1.1.**

Few families are composed of members speaking different native language. Meanwhile, the specific data is hard to derive. So, we ignore the influence of the multi-composed families and assume that children all take their parents' native language as their own native language.

*More detailed assumptions will be listed if needed.*

### 3. Notation

|           |   |
|-----------|---|
| $A_i$     | The percentage of people in country $i$ speak a second language.  |
| $P_{ij}$  | The possibility of people in country $i$ speak a second language $j$ . If language $j$ is the native language of this country, $P_{ij} = 0$ . |
| $NONLS_i$ | Number of native language speakers in country $i$ .   |
| $NOSLS_i$ | Number of second language speakers in country $i$ .   |
| $TNOS_i$  | Total number of speakers in country $i$ .   |

## 4. Model

### 4.1 Language Development Model

To discuss the changes of the top-ten language list, we analyze 16 currently most spoken languages. We gather the characteristics of language distribution of different countries to produce a global language distribution. To reduce the amount of calculation, we only consider countries with a population over 10 million, since they have contributed about 92% of total population on Earth as shown in Appendix 1. To simplify our model, we suggest the language with the most speakers represent the native language of a certain country. For each country, we acquire its population data from 2007 to 2017 on the United Nation Database [1], and obtain its Net Birth Rate to calculate its newly-born population every year. Meanwhile, the percentage of different language speakers is acquired [1,2] so that we can know the current language structure of the country. Through careful analysis of every country we take into account, we seek to obtain the language distribution all over the world.

#### 4.1.1 Judgment Process

We assumed that the existing people won't make a change on their native and second language, and the newly-born population all take their parents' native language as their own native language. Therefore, the numerical development of a certain language is completely determined by the newly increased population. However, not all of them will speak a second language, and which language they will take as their second language varies for economic, social, political and other reasons. So, we divide the judgment process into two steps – first to judge whether one will speak a second language, and then which language he/she will take. The judgment process of our Language Development Model is shown in Figure 1.

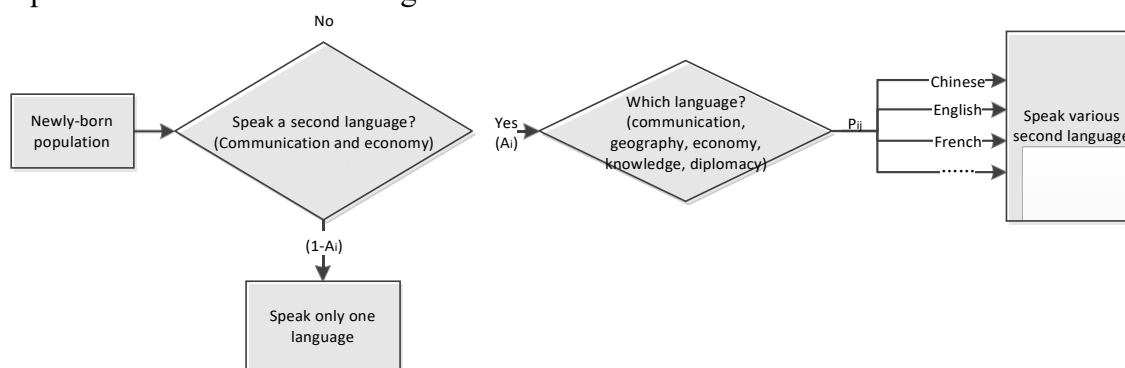


Figure 1 Judgment process of Language Development Model

To quantify various reasons that affect the second language, we first take communicative and economic factors to determine if one speaks a second language in our analysis. Both two factors weigh 0.5 when calculating the possibility.

**Communicative factor** is measured by diversity, which equals to number of language that is spoken by over 100,000 people to 16 (the number of commonly used language we discuss in the paper). For example, if there are 10 languages spoken by over 100,000 people in that country, the score of communicative factor equals to  $10/16$  (0.625). So, the higher the diversity of language, the higher the percentage  $A_i$  for newly-born generation to speak a second language.

However, when considering **economic factors**, the economic development level may not have positive correlation to the percentage  $A_i$ . As far as we are concerned, people in developed countries may lack the motivation to study a second language although they have sufficient resources, while people in backward countries lack resources to learn. Therefore, people in developing countries who have moderate resources and highest passion have the highest likelihood to study. The level of economic development is divided into 3 groups according to the GDP per capita of the country. In the meantime, GDP per capita increases at different rate for countries in different development level, which is listed in

Table 1 below. Taking all these factors considered, we give a covering possibility of speaking a second language for every country we discussed above.

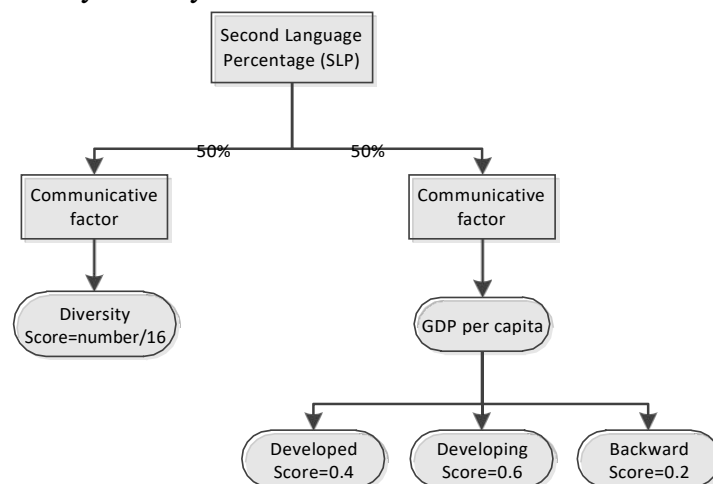


Figure 2 Principle to determine whether one speaks a second language

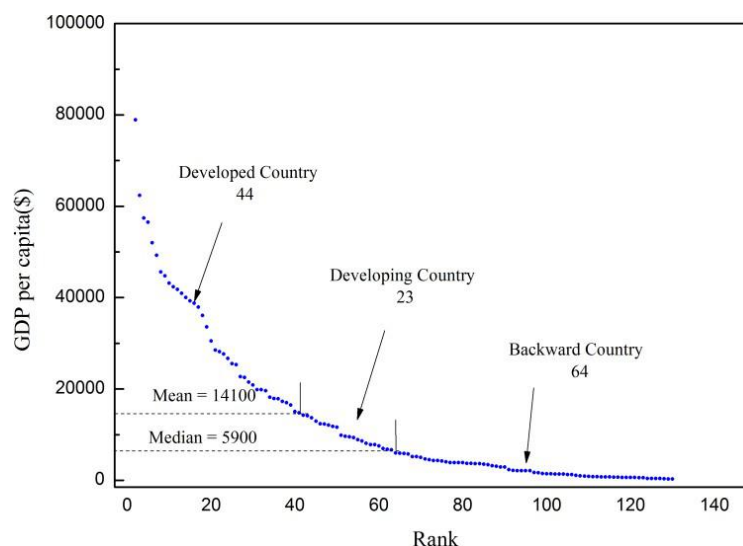


Figure 3 Classification of countries according to GDP per capita

Table 1 Increasing rate of GDP per capita for countries in different development level

| Countries | Developed | Developing | Backward |
|-----------|-----------|------------|----------|
| Rate      | +2%       | +7%        | +3%      |

As for the choice of the second language, we take five factors into consideration – communication[3], geography, economy, knowledge[4] and diplomacy[5]. Each of the five factors contains some measurable parameters, for example, number of countries that speak the language (NOCS) and number of people that speak the language (NOPS) in geography section, GDP and number of top 10 economics in economy section. Among these parameters, some are global and have the same value for every country, such as NOCS and NOPS, while others are internal and differ from country to country, such as local GDP and domestic language distribution. This means people in different countries have different possibility to speak the same second language ( $P_{i1j1} \neq P_{i2j1}$ ), and people in the same country have different possibility to speak different second language ( $P_{i1j1} \neq P_{i1j2}$ ).

To determine the weights of each of the five sections, we use Analytical Hierarchy Process (AHP) to determine the weights. This method helps to convert our subjective preference to measurable weights to evaluate the importance of different factors.

We first build a 5\*5 comparison matrix, each element of which shows the extent of preference between factor i and j. Since daily communication is important for people, and studying a second language usually have a close relationship to business and trade, we give them special preference, and obtain the following matrix.

|     | Com | Geo | Eco | Kno | Dip |
|-----|-----|-----|-----|-----|-----|
| Com | 1   | 2   | 1   | 5   | 6   |
| Geo | 1   | 1/2 | 1   | 3   | 5   |
| Eco | 1   | 2   | 1   | 5   | 6   |
| Kno | 1/5 | 1/3 | 1/5 | 1   | 2   |
| Dip | 1/6 | 1/5 | 1/6 | 1/2 | 1   |

Calculating the maximum eigenvalue and normalizing the corresponding eigenvector, we obtain the weights of each section, and divide them equally for the measurable subsections. The sections we consider together with their weight is shown in Figure 4 below.

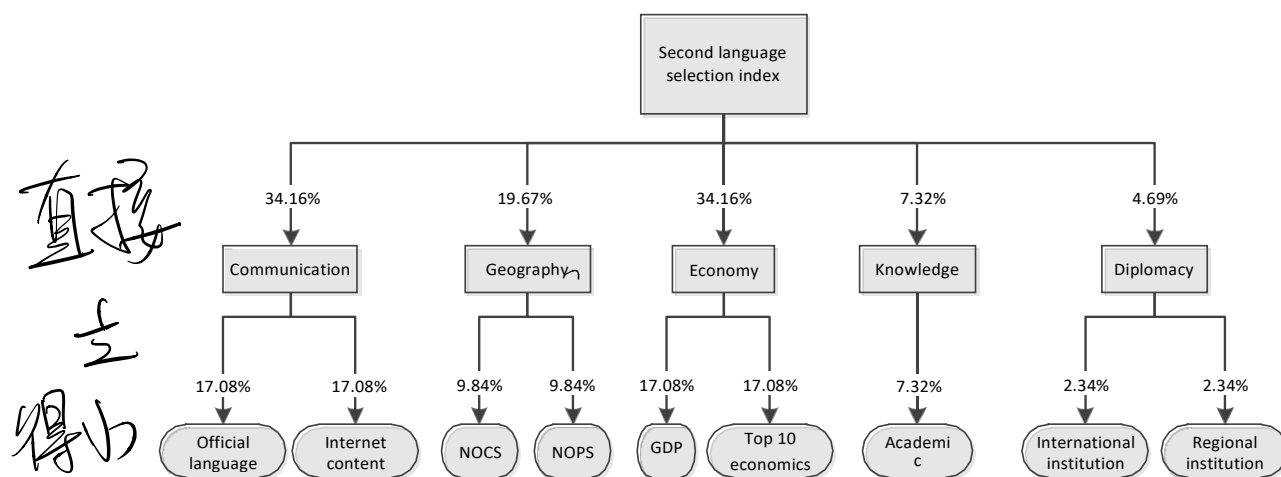


Figure 4 Sections and their weight when determining which second language to speak

英(1, 2, 3, 4, 5) × 权 = 得分



To test the consistency of the matrix, we calculate the Consistency Ratio (CR), which is defined as the ratio of Consistency Index (CI) to Average Random Consistency Index (RI). Since  $n = 5$ ,  $RI = 1.12$ , and  $CI = \frac{\lambda_{max} - n}{n - 1} = 0.0115$ , we get  $CI = 0.0103 < 0.1$ , so the consistency of the matrix is confirmed.

#### 4.1.2 Stochastic Simulation Process

At the initial phase, we simply simulate the total number of speakers ( $TNOS_i$ ) to simplify our model. Meanwhile, we take the dominant language to represent the native language of a certain country, and ignore the influence of migration. Therefore, the change of  $TNOS_i$  is simply determined by the newly-born generation. When simulating the model on Python, we first calculate the product of Second Language Percentage (SLP) and the newly-born population of the country to determine the increased number of second language learner. Then we assign a random value between 0 and 1 to every 10,000 second language learners to determine which second language to learn. If the random value falls within the interval of a certain language, all people in the group are regarded to study the specific language.

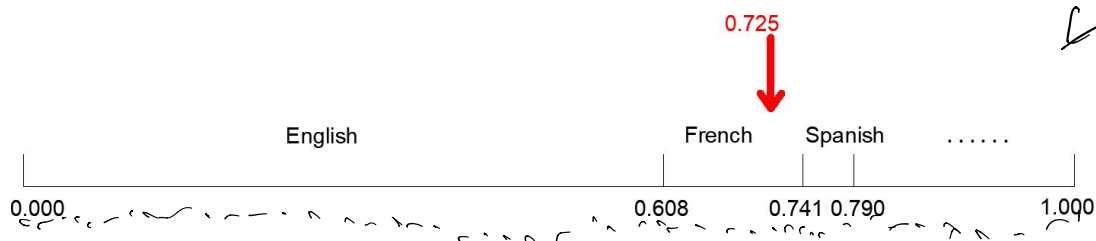


Figure 5 Determination of which second language to speak

The first simulation result shows that moderate changes occur to the top ten list of total number of speakers. Two languages – Portuguese and Russian, are replaced by German and Japanese. The rank of the eight languages listed both in 2017 and 2067 also shows a slight difference.

As for the total number of speakers, most of the languages we analyze gain growth. The number of German speakers increases the fastest, up to 805.11%. The number of French and Japanese speakers also increases by over 500%. This is mainly because countries which speak these languages have high GDP, so people around the world tend to study these languages for economic reasons. Meanwhile, the original population base of these languages is not so high as Chinese/English speaking countries. The combination of the two aspects leads to this stunning result. If the threshold is reduced to 100%, up to 6 languages are included. But the language which develops the worst, the Russian, only record a -0.51% drop. This is due to the negative Net Birth Rate in Russia and few countries speaking Russian, which lead to decrease in domestic population and lack of new learners for communicative factors.

Table 2 First simulation result

| Language   | Rank in 2017 | Rank in 2067 | Population in 2017 | Population in 2067 | Increased rate (%) | Increased number |
|------------|--------------|--------------|--------------------|--------------------|--------------------|------------------|
| English    | 1            | 1            | 1255857851         | 3172666602         | 152.63             | 1916808751       |
| Chinese    | 2            | 2            | 1096961483         | 2278106986         | 107.67             | 1181145503       |
| Spanish    | 3            | 3            | 520236718          | 1482532068         | 184.97             | 962295350        |
| Hindi      | 4            | 7            | 449573687          | 480031690          | 6.77               | 30458003         |
| Arabic     | 5            | 8            | 334611664          | 345470847          | 3.25               | 10859183         |
| Malay      | 6            | 9            | 292434196          | 296957189          | 1.55               | 4522993          |
| Bengali    | 7            | 10           | 246276800          | 248034947          | 0.71               | 1758147          |
| French     | 8            | 4            | 242897420          | 1471076958         | 505.64             | 1228179538       |
| Portuguese | 9            | 11           | 229370143          | 238969091          | 4.18               | 9598948          |
| Russian    | 10           | 12           | 207790825          | 206735561          | -0.51              | -1055264         |
| Hausa      | 11           | 13           | 150021627          | 149734135          | -0.19              | -287492          |
| Punjabi    | 12           | 14           | 147803740          | 149154483          | 0.91               | 1350743          |
| Japanese   | 13           | 6            | 133972338          | 1135657916         | 747.68             | 1001685578       |
| German     | 14           | 5            | 136642027          | 1236759996         | 805.11             | 1100117969       |
| Persian    | 15           | 15           | 119207805          | 125462681          | 5.25               | 6254876          |
| Urdu       | 16           | 16           | 69081313           | 69250144           | 0.24               | 168831           |

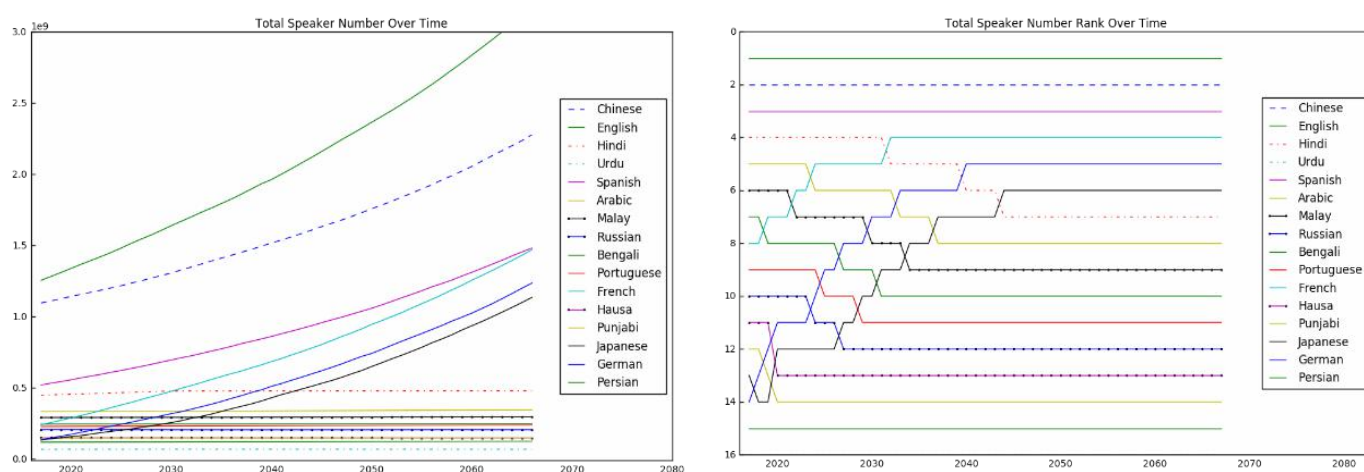


Figure 6 Change of total number and rank of the 16 languages over 50 years



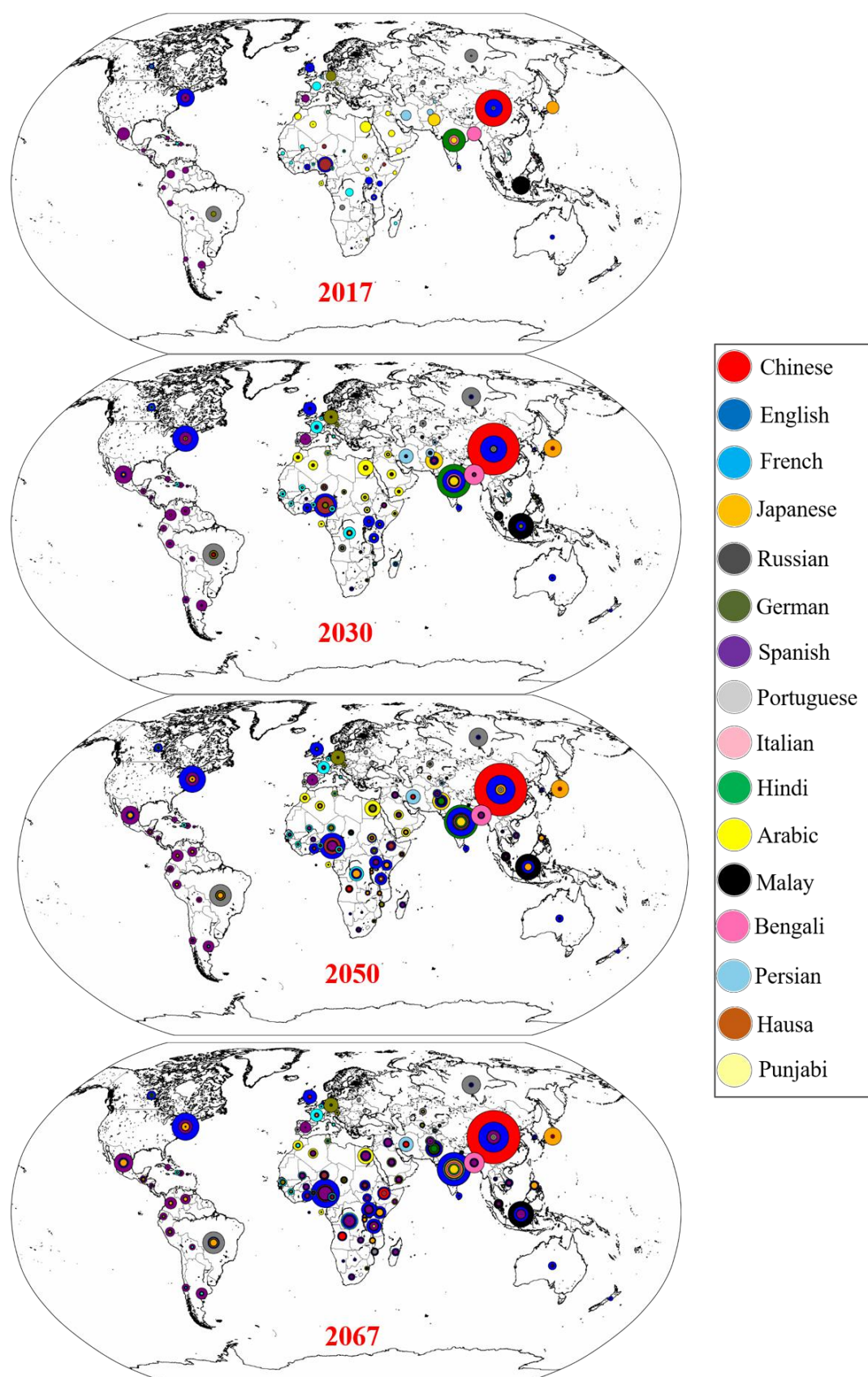


Figure 7 Change of language distribution around the world

To improve our model, we consider number of native language speakers ( $NONLS_i$ ) and number of second language speakers ( $NOSLS_i$ ) separately in the second phase. Since the total number is composed by native and second language parts,  $TNOS_i = NONLS_i + NOSLS_i$ , we just need to simulate the number of native speakers, and get the number of second language speakers

by subtraction. We get the following results:

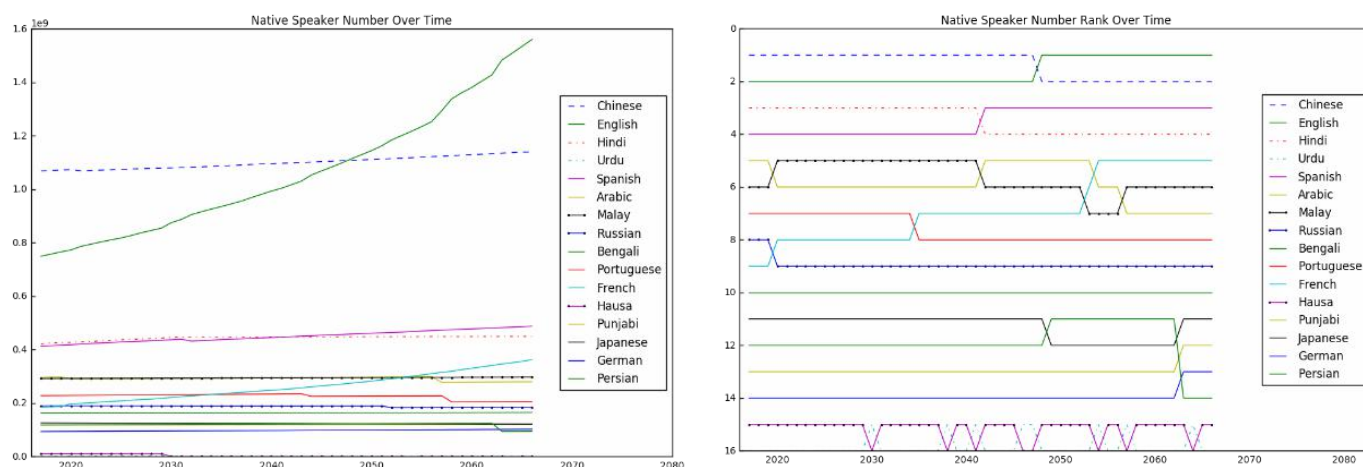


Figure 8 Change of number of native speakers and rank of 16 languages over 50 years

Compared with the result of total speakers, we can see that the initial number of native Chinese speakers is more than that of English. However, native English speakers increase much faster than Chinese ones, and surpass Chinese in 2049. The number of other native language speakers also have a small difference from the total number. Meanwhile, we can see that the curve of native speakers is not so smooth as the total speakers one. This is the result of our simplification that we only take the dominant language to represent the native language of a country, so there may be sudden growth or drop of the number when the prevailing language of a country changes.

To better simulate the reality situation, we further consider the influence of migration. We assume that the moment the migration happens, all the immigrants instantly change their native language from their original one to the new one. The emigrant country subtracts the same number of native language speakers as emigrant number, and immigrant country adds the number to their native language speakers. Meanwhile, immigrants' original native language is counted into the second language of the new country.

Taking the influence of migration into account, little change occurs in the total number and native language speaker number of all these 16 languages. The development trend and speed are all similar to the former one without consideration of migration. This is attributed to the inconspicuous rate of migration, which is approximately one tenth of the average Net Birth Rate. Simulation result is shown in the following figure:

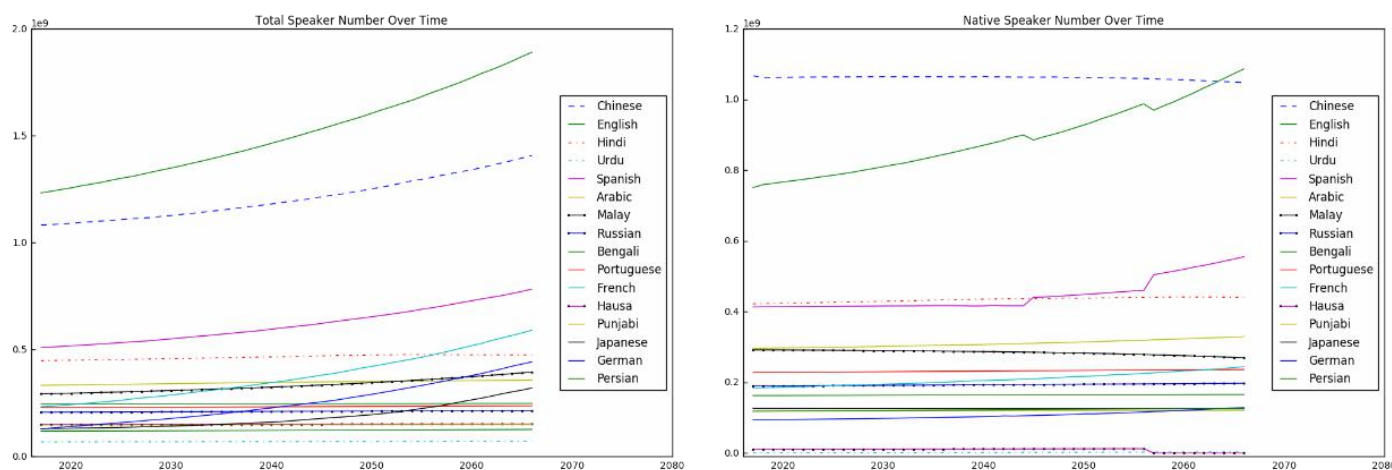


Figure 9 Simulation result when take migration into account

## 4.2 Location Determination Model

### 4.2.1 Solution Oriented by Language Structure

To determine the location of the new offices, we build a Location Determination Model, using the results obtained from the Stochastic Language Development Model. Since international branch offices are usually set in well developed country, we simply focus on countries with GDP over 100 billion dollars, which helps reduce the number of country to 52. In the meantime, the capital of a country is usually the economic, political and transportation center, so we adopt the capital as the only candidate city for each country.

For these 52 countries (or you can say, their capitals), we employ Algorithms for Fuzzy Clustering (AFFZ) to separate them into 6 groups according to their similarity in language structure[6]. For each country, it generates a vector, each element of which stands for the total number of people speaking the specific language. The correlation between every two countries can be calculated through Angle Cosine Method:

$$r_{ij} = \frac{|\sum_{k=1}^m x_{ik} \cdot x_{jk}|}{\sqrt{\sum_{k=1}^m x_{ik}^2 \cdot \sum_{k=1}^m x_{jk}^2}}$$

With the correlation value, we create a Fuzzy Similar Matrix, and then generate a Fuzzy Equivalent Matrix to divide these 52 countries into 6 groups. For each group, we choose 1-3 languages to represent the characteristics of it.

The purpose of establishing an international new office is to get tight connection with economically developed countries and regions, so there are two objectives that we are concerned about most, namely the distance from the candidate country (or capital) to other countries, and the GDP distribution in the group. We hope that the office is built somewhere close to the well-developed countries, and we get the following expression for country  $j$

$$\sum_{i \neq j} \frac{(GDP_i)^2}{d_{ij}} + \frac{(GDP_j)^2}{\frac{1}{2} \min_{i \neq j} d_{ij}}$$

In the expression, subscript  $i$  refers to country  $i$ , and  $d_{ij}$  refers to the distance between country  $i$  and  $j$ . We square the GDP since we think that GDP is a more important factor compared with distance. The first section of the expression represents the effect of distance as well as GDP of other countries. When considering domestic condition, we also use the data of GDP and distance to keep the dimension coherent to the former. However, the distance can't be 0, so we replace it with half of the distance from its closest country, which makes the value a minimum and not to be 0.

The larger the value of the expression, the better a country to set the office in. So, we need to search for the maximum value in each group to determine the location. That means, the objective function is

$$\max(\sum_{i \neq j} \frac{(GDP_i)^2}{d_{ij}} + \frac{(GDP_j)^2}{\frac{1}{2} \min_{i \neq j} d_{ij}})$$

Running the solution program on Matlab using the current (2017) language structure of each country, we obtain the members of each group as the following table shows:

*Table 3 Current grouping result (solution 1)*

| Group | Language  | Countries  | Number |
|-------|---|--|--------|
| 1     | <b>Arabic</b>                                     | Morocco, Algeria, Egypt, Iraq, Saudi Arabia  | 5      |
| 2     | <b>Spanish<br/>+<br/>French<br/>+<br/>Russian</b> | South Africa, Nigeria, Armenia, Kazakhstan, Kyrgyzstan, Vietnam, Singapore, Philippine, Thailand, South Korea, Japan, Romania, Greece, Czechia, Finland, Austria, Belgium, Poland, Switzerland, Netherland, Spain, Russia, Italy, France, Britain, Germany, Mexico, Puerto Rica, Canada, New Zealand, Australia, Peru, Chile, Colombia, Venezuela, Argentina, China, America | 38     |
| 3     | <b>Bengali</b>                                    | Bangladesh   | 1      |
| 4     | <b>Malay</b>                                      | Malaysia, Indonesia  | 2      |
| 5     | <b>Hindustani</b>                                 | India, Pakistan  | 2      |
| 6     | <b>Portuguese</b>                                 | Portugal, Brazil   | 2      |

Based on the latitude and longitude data of each country, we calculate the distance among them with simple Matlab program. Then we can get the value of the criterion expression of each country, therefore determine the countries for new offices. Final result of locations and the languages used in each office is shown in Table 4.

*Table 4 Current locations and languages for each office*

| Country<br>(Capital) | <b>Saudi Arabia<br/>(Riyadh)</b> | <b>Germany<br/>(Berlin)</b>   | <b>Bangladesh<br/>(Dhaka)</b> | <b>Indonesia<br/>(Jakarta)</b> | <b>India<br/>(New Delhi)</b> | <b>Brazil<br/>(Brasilia)</b> |
|----------------------|----------------------------------|-------------------------------|-------------------------------|--------------------------------|------------------------------|------------------------------|
| Language             | Arabic                           | Spanish,<br>French,<br>Russia | Bengali                       | Malay                          | Hindustani                   | Portuguese                   |

If we substitute the data of language structure and GDP with the future one in 2067 (GDP is regarded to increase at a constant speed as shown in

Table 1), the grouping result and the locations as well as spoken languages will change as Table 5 and Table 6 show below:

*Table 5 Future grouping result*

| Group | Language   | Countries  | Number |
|-------|--|--|--------|
| 1     | <b>Arabic</b>                                      | Morocco, Algeria, Egypt, Iraq, Saudi Arabia  | 5      |
| 2     | <b>Hindustani<br/>+<br/>Malay<br/>+<br/>German</b> | Brazil, Russia, Poland, India, Germany, Malaysia, Britain, Kazakhstan, Indonesia, Canada, South Korea, Australia, Netherland, Iran, Thailand, Switzerland, Nigeria, Philippine, South Africa, Pakistan, Austria, Vietnam, Singapore, Finland, Czechia, New Zealand, Botswana, China, America | 29     |
| 3     | <b>Bengali</b>                                     | Bangladesh   | 1      |
| 4     | <b>Spanish</b>                                     | Mexico, Argentina, Venezuela, Chile, Peru, Romania, Italy, Spain, Ecuador, Cuba, Dominican   | 13     |



| # Team |                 |                              |  | Page 13 of 23 |
|--------|-----------------|------------------------------|--|---------------|
|        |                 | Republic, Colombia, Portugal |  |               |
| 5      | <b>Japanese</b> | Japan                        |  | 1             |
| 6      | <b>French</b>   | France, Belgium              |  | 2             |

Table 6 Future locations and languages for each office

| Country<br>(Capital) | Saudi Arabia<br>(Riyadh) | Poland<br>(Warsaw)              | Bangladesh<br>(Dhaka) | Japan<br>(Tokyo) | France<br>(Paris) | Argentina<br>(Buenos Aires) |
|----------------------|--------------------------|---------------------------------|-----------------------|------------------|-------------------|-----------------------------|
| Language             | Arabic                   | Hindustani,<br>Malay,<br>German | Bengali               | Japanese         | French            | Portuguese                  |

#### 4.2.2 Solution Oriented by Geographic Condition

As we can see, the result produced by solution oriented by language structure can hardly be regarded as the ideal locations for new office. In fact, although relatively little languages are considered such as Arabic and Bengali and it is easy to determine what languages to speak in each new office, the locations are unsatisfactory from the economic point of view. This can be stem from the huge difference of the 6 groups, one of which contains 38 countries spreading on almost every continent, covering nearly 75% of the total countries we discuss. The remaining 14 countries are all relatively small or medium countries, and have mediocre economic level, which are not suitable locations for international business.

Since then, we shift our way of thinking, and first **decide how many offices** to build on each continent (totally 6 worldwide) according to the economic development. For continents **with n ( $n > 1$ ) offices**, we divide them into n parts based on the geographic location of each country, applying P Center Method while grouping. That means, if there are two offices in Europe, we randomly take two countries of it, and **sum up the smaller value of the distance** between another country and these two. After traversing all cases, we select two countries with the minimum value as the reference countries, marked as A, B. Then we add other countries into the group of country A or B on the basis of their distance from these countries. For example, another country C has shorter distance to country B than country A, then it is added to the second group.

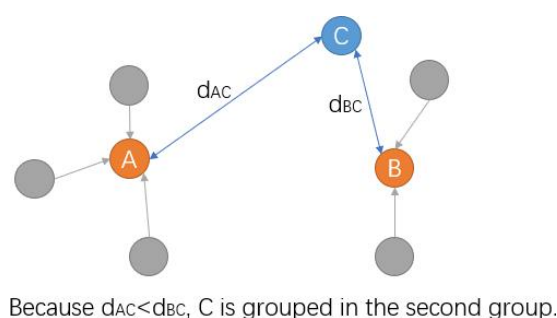
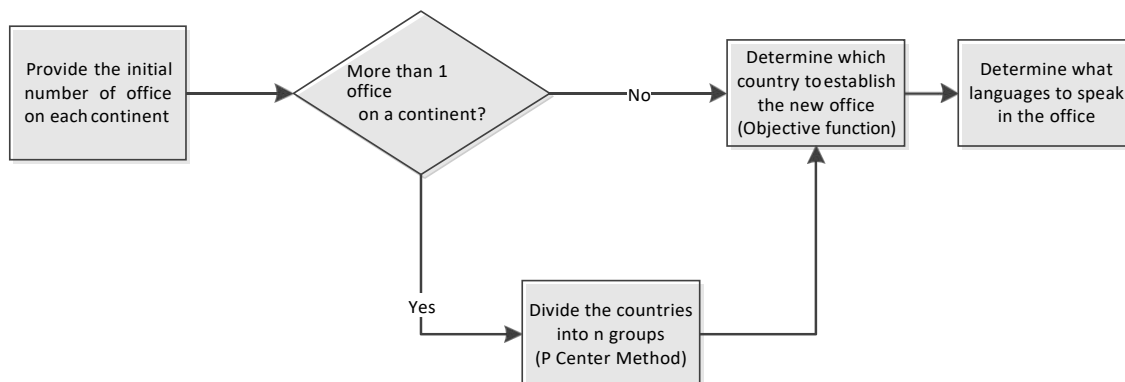


Figure 10 The grouping process

After the grouping process, we use the same criterion expression as the solution oriented by language structure to determine the optimal office location for each group. Algorithms for Fuzzy Clustering (AFFZ) is applied to finally decide what language to speak in the office. We observe the characteristics of language structure of clustered group with most countries, and choose 1-3

languages that have great amount of speakers.



*Figure 11 Process of solution oriented*

When simulating the solution on Matlab program, we take the original two offices in Shanghai and New York into consideration since they also contribute to connection with other countries around them, which may influence the location selection. That means, we suggest originally there is no office anywhere on Earth, and we need to distribute 8 offices around the world, and see if the results include companies in China and the United States. The number of offices distributed around the world is initially set as the following table:

*Table 7 Initial distribution of office number*

| Continent | Asia | Europe | North America | South America | Oceania | Africa |
|-----------|------|--------|---------------|---------------|---------|--------|
| Number    | 2    | 2      | 1             | 1             | 1       | 1      |

For Asia and Europe, two offices are set, so P Center Method is applied to separate their countries into two groups. For other continents, only one office is set, so the grouping process is skipped. Russia is regarded as an Asian country because geographically most of its territory locates in Asia.

We discuss the current and future situation separately. The result of the 8 groups for current situation is shown as following:

*Table 8 Current grouping result (Solution 2)*

| Group | Continent     | Members     |             |              |             |             |
|-------|---------------|-------------|-------------|--------------|-------------|-------------|
| 1     | Asia          | Vietnam     | Bangladesh  | Singapore    | Thailand    | South Korea |
|       |               | Japan       | China       | Malaysia     | Philippines | Indonesia   |
| 2     | Asia          | India       | Pakistan    | Armenia      | Kazakhstan  | Kyrgyzstan  |
|       |               | Iraq        | Iran        | Saudi Arabia | Russia      |             |
| 3     | Europe        | Hungary     | Greece      | Austria      | Poland      | Romania     |
|       |               | Czechia     | Finland     | Italy        |             |             |
| 4     | Europe        | Belgium     | Switzerland | Netherlands  | France      | Britain     |
|       |               | Germany     | Portugal    | Spain        |             |             |
| 5     | North America | Puerto Rico | Mexico      | Canada       | America     |             |
| 6     | South America | Peru        | Chile       | Colombia     | Venezuela   | Argentina   |
|       |               | Brazil      |             |              |             |             |
| 7     | Oceania       | New Zealand | Australia   |              |             |             |
| 8     | Africa        | Morocco     | Algeria     | South Africa | Egypt       | Nigeria     |



Using the objective function mentioned above to evaluate the suitability of each country in a group, we obtain the optimal location for the new offices in 2017. The result also validate that China and the United States of America are ideal locations to set offices in. Figure 12 distinctly shows the location of offices and corresponding relative members in the same group on the map of the world.

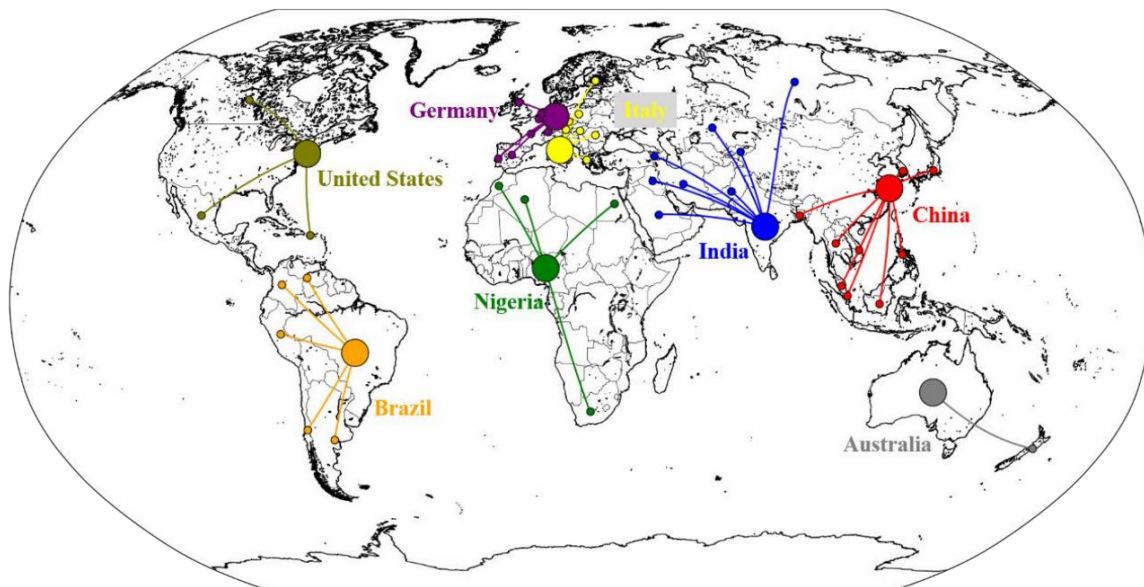


Figure 12 Location of offices worldwide in 2017

The languages spoken in each country are shown in the table below:

Table 9 Languages spoken in each office in 2017

| Country<br>(Capital) | Nigeria<br>(Abuja) | India<br>(New<br>Delhi)            | China<br>(Shanghai)  | Italy<br>(Rome) | Germany<br>(Berlin)           | America<br>(New<br>York) | Brazil<br>(Brasilia) | Australia<br>(Canberra) |
|----------------------|--------------------|------------------------------------|----------------------|-----------------|-------------------------------|--------------------------|----------------------|-------------------------|
| Language             | Arabic             | Russian,<br>Hindistani,<br>Punjabi | Chinese,<br>Japanese | German          | German,<br>French,<br>Spanish | Spanish,<br>French       | Spanish              | Spanish                 |

If the data is replaced by the one in 2067, Italy and India are replaced by Poland and Russia. China and America still exist in the list, once again validating their suitability to be the location for international offices. Meanwhile, small changes occur in languages spoken because of different develop speed of different languages. The grouping result, the final locations of offices and the languages spoken are shown below.

Table 10 Future grouping result (solution 2)

| Group | Continent     | Members     |              |              |             |                    |
|-------|---------------|-------------|--------------|--------------|-------------|--------------------|
| 1     | Asia          | Vietnam     | Bangladesh   | Singapore    | Thailand    | South Korea        |
|       |               | Japan       | China        | Malaysia     | Philippines | Indonesia          |
| 2     | Asia          | India       | Pakistan     | Saudi Arabia | Kazakhstan  | Russia             |
|       |               | Iraq        | Iran         |              |             |                    |
| 3     | Europe        | Hungary     | Greece       | Austria      | Poland      | Romania            |
|       |               | Czechia     | Finland      | Italy        | Croatia     |                    |
| 4     | Europe        | Belgium     | Switzerland  | Netherlands  | France      | Britain            |
|       |               | Germany     | Portugal     | Spain        |             |                    |
| 5     | North America | Cuba        | Mexico       | Canada       | America     | Dominican Republic |
| 6     | South America | Peru        | Chile        | Colombia     | Venezuela   | Argentina          |
|       |               | Brazil      | Ecuador      |              |             |                    |
| 7     | Oceania       | New Zealand | Australia    |              |             |                    |
| 8     | Africa        | Algeria     | South Africa | Egypt        | Nigeria     |                    |

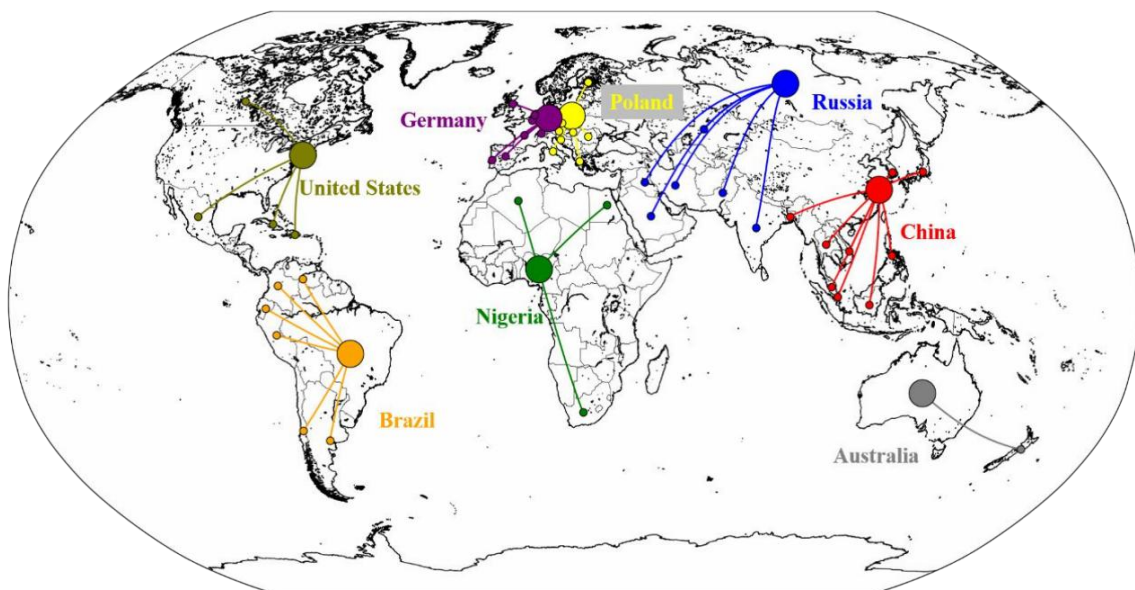


Figure 13 Location of offices worldwide in 2067

Table 11 Languages spoken in each office in 2067

| Country (Capital) | Nigeria (Abuja) | Russia (Moscow)     | China (Shanghai)  | Poland (Warsaw) | Germany (Berlin)        | America (New York) | Brazil (Brasilia) | Australia (Canberra) |
|-------------------|-----------------|---------------------|-------------------|-----------------|-------------------------|--------------------|-------------------|----------------------|
| Language          | Arabic          | Russian, Hindistani | Chinese, Japanese | German          | German, French, Spanish | Spanish            | Spanish           | Malay                |

## 5. Sensitivity Analysis

### 5.1 Sensitivity Analysis on Language Development Model

In the Language Development Model, the Net Birth Rate, GDP growth rate and migration rate will have an impact on the results. Therefore, it is necessary to analyze the impact of these three factors on the simulation results.

#### 5.1.1 Net Birth Rate Impact

The Net Birth Rate in each country is the main motivation behind the language development, here we reduce and increase the net birth rate of all countries by 20% and 10%, and use the Language Development Model to calculate the total number of 16 languages speakers after 50 years. The relative changes are shown in the following table.

*Table 12 Analytical result of Net Birth Rate impact*

| Change of Net Birth Rate | -20%    | -10%    | +10%   | +20%   |
|--------------------------|---------|---------|--------|--------|
| <b>Chinese</b>           | -13.60% | -6.50%  | 10.53% | 20.22% |
| <b>English</b>           | -13.84% | -5.41%  | 13.07% | 24.76% |
| <b>Hindi</b>             | -0.05%  | -0.55%  | 0.85%  | 0.73%  |
| <b>Urdu</b>              | 0.75%   | -0.84%  | -0.16% | -1.03% |
| <b>Spanish</b>           | -18.45% | -9.99%  | 10.34% | 23.82% |
| <b>Arabic</b>            | 3.20%   | -3.05%  | 2.41%  | 2.30%  |
| <b>Malay</b>             | -1.56%  | -1.53%  | 1.53%  | 1.50%  |
| <b>Russian</b>           | -1.89%  | -1.91%  | -1.01% | -0.78% |
| <b>Bengali</b>           | -0.72%  | -0.76%  | 0.73%  | 0.74%  |
| <b>Portuguese</b>        | -4.66%  | -4.42%  | 4.29%  | 3.87%  |
| <b>French</b>            | -24.35% | -13.02% | 14.00% | 27.69% |
| <b>Hausa</b>             | -0.26%  | -0.04%  | 0.02%  | -0.19% |
| <b>Punjabi</b>           | -1.28%  | 0.88%   | 1.07%  | 0.93%  |
| <b>Japanese</b>          | -28.38% | -13.96% | 17.91% | 34.08% |
| <b>German</b>            | -25.32% | -14.07% | 16.50% | 36.66% |
| <b>Persian</b>           | -5.90%  | -5.69%  | 5.39%  | 5.27%  |

As can be seen from the above table, for most languages, the number of language speakers is positively correlated with the Net Birth Rate. In other words, as the Net Birth Rate rises, the number of language speakers increases, and as the Net Birth Rate decreases, the number of language speakers drops. This result is consistent with common sense. From this perspective, we can say that this model accurately simulates the reality situation. In addition, all the changes in this table are less than 40%. Therefore, the model can be considered as a robust one.

#### 5.1.2 GDP Growth Rate Impact

The growth rate of GDP used in the Language Development Model was 2% in developed countries, 7% in developing countries and 3% in backward countries. Now we change these three numbers by -50%, -10%, +10%, +100%, and use the Language Development Model to calculate the total number of 16 languages speakers after 50 years. The relative changes are shown in the

following table.

*Table 13 Analytical result of GDP Growth Rate impact*

| GDP increase rate | -50%   | -10%   | +10%   | +100%  |
|-------------------|--------|--------|--------|--------|
| <b>Chinese</b>    | -0.51% | -0.49% | 1.45%  | 2.49%  |
| <b>English</b>    | -2.88% | -2.15% | 3.62%  | 1.52%  |
| <b>Hindi</b>      | 0.32%  | 0.69%  | 0.24%  | 1.40%  |
| <b>Urdu</b>       | -0.28% | -0.30% | -0.07% | 0.05%  |
| <b>Spanish</b>    | -0.20% | -1.44% | -2.20% | 0.99%  |
| <b>Arabic</b>     | -2.47% | -2.71% | 2.82%  | 2.16%  |
| <b>Malay</b>      | -1.58% | -1.58% | 1.58%  | 1.58%  |
| <b>Russian</b>    | -0.62% | -0.45% | -0.65% | 0.84%  |
| <b>Bengali</b>    | -0.75% | -0.75% | 0.75%  | 0.75%  |
| <b>Portuguese</b> | -4.32% | -4.34% | 4.30%  | 4.39%  |
| <b>French</b>     | -1.64% | -1.92% | 0.62%  | 0.27%  |
| <b>Hausa</b>      | -0.13% | -0.13% | -0.13% | -0.13% |
| <b>Punjabi</b>    | -1.96% | -0.73% | 0.85%  | 1.56%  |
| <b>Japanese</b>   | -1.18% | -0.80% | 1.68%  | 3.66%  |
| <b>German</b>     | -2.38% | 0.38%  | 0.30%  | 1.71%  |
| <b>Persian</b>    | 5.50%  | 5.58%  | 5.58%  | 5.50%  |

As can be seen from the above table, the increase in GDP growth rate will cause an increase in the number of most languages. However, even in the case of GDP growth rate increases by 100%, the total population of languages will not change much (the maximum is only 3.66%). Therefore, the change in the GDP growth rate has little effect on the result of the population after 50 years.

### 5.1.3 Migration Rate Impact

By increasing and decreasing the migration rate in each country, we calculate the total population change of each language after 50 years.

*Table 14 Analytical result of Migration Rate impact*

| Migration Rate    | -20%   | -10%   | +10%   | +20%    |
|-------------------|--------|--------|--------|---------|
| <b>Chinese</b>    | 3.16%  | -0.81% | -3.51% | -8.30%  |
| <b>English</b>    | -4.23% | -2.18% | 3.40%  | 5.27%   |
| <b>Hindi</b>      | 1.14%  | 0.00%  | -5.20% | -6.86%  |
| <b>Urdu</b>       | 0.00%  | 0.00%  | 0.19%  | 0.00%   |
| <b>Spanish</b>    | -0.12% | -0.08% | 1.10%  | 1.42%   |
| <b>Arabic</b>     | 0.15%  | 0.40%  | -1.34% | -1.77%  |
| <b>Malay</b>      | -3.80% | -6.94% | -9.11% | -23.48% |
| <b>Russian</b>    | 0.20%  | -0.01% | -0.40% | -0.34%  |
| <b>Bengali</b>    | 0.08%  | 0.06%  | 0.03%  | -0.01%  |
| <b>Portuguese</b> | -0.17% | -0.15% | -0.07% | -0.05%  |
| <b>French</b>     | -1.00% | 0.09%  | -0.60% | -0.23%  |

|                 |        |        |        |        |
|-----------------|--------|--------|--------|--------|
| <b>Hausa</b>    | -0.11% | -0.15% | -0.13% | -0.07% |
| <b>Punjabi</b>  | 0.06%  | -0.02% | 0.04%  | 0.01%  |
| <b>Japanese</b> | -0.10% | -1.44% | 8.51%  | 3.59%  |
| <b>German</b>   | -0.51% | -0.32% | -1.12% | 4.30%  |
| <b>Persian</b>  | -0.68% | -0.44% | -0.59% | 0.17%  |

From this table, we can see that the number of Chinese, Hindi and Malay speakers are negatively correlated with migration rate, while the total number of speakers of English and Japanese are positive correlated with it. This is because people speaking Chinese, Malay and Hindi have stronger tendency to migrate and lots of them choose English-speaking countries as their second hometown.

## 5.2 Sensitivity Analysis on Location Determination Model

As is shown in 4.2, three factors contribute to the final locations for the new offices – geographic position, language structure and economic condition measured by GDP.

In our model, the impact of a country's geographic position is expressed by distance. As the position of a country is not likely to change over time, we do not pay much attention to the sensitivity of this factor. Language structure, which is affected by net birth rate, migration and GDP growth, plays a significant role when deciding what language should be spoken in the new offices, but it is less important in the position-determining step. Consequently, in the sensitivity analysis section, we mainly focus on the fluctuation of GDP growth and its influences.

Similar as what we have done in section 5.1.2, we again apply a -50%, -10%, +10% and +100% variation to the GDP growth rate, and record the recommending countries for new offices under each circumstance, respectively. We spot notable changes.

*Table 15 Sensitivity Analysis of Location Determination Model*

|                           | Group members Variation | Recommended Country | Group members Variation | Recommended Country | Group members Variation | Recommended Country | Group members Variation | Recommended Country |
|---------------------------|-------------------------|---------------------|-------------------------|---------------------|-------------------------|---------------------|-------------------------|---------------------|
| GDP growth rate variation | -50%                    |                     | -10%                    |                     | +10%                    |                     | +100%                   |                     |
| Africa                    | +0                      | Nigeria             | +0                      | Nigeria             | +2                      | South Africa        | +4                      | Botswana            |
| Asia                      | +0                      | Russia              | +0                      | Russia              | +0                      | Russia              | +0                      | Russia              |
|                           | +0                      | China               | +0                      | China               | +0                      | China               | +0                      | China               |
| Europe                    | +0                      | Germany             | +0                      | Germany             | +0                      | Germany             | -2                      | Germany             |
|                           | +0                      | Poland              | +0                      | Poland              | -1                      | Poland              | -1                      | Poland              |
| North America             | +0                      | USA                 | +0                      | USA                 | +0                      | Mexico              | +0                      | Mexico              |
| Oceanic                   | +0                      | Australia           | +0                      | Australia           | -1                      | Australia           | -1                      | Australia           |
| South America             | +0                      | Brazil              | +0                      | Brazil              | +0                      | Brazil              | +0                      | Brazil              |

From the table above we can conclude that:

- If the GDP growth rate decreases, neither the group members nor the recommended country changes.
- If the GDP growth rate increases, there tends to be more group members for Africa and less members for Europe and Oceanic. Meanwhile, the recommended countries alter.

This phenomenon can be explained by the difference of GDP growing speed between developing and developed countries. When GDP growth rate increases, the difference broadens, meaning that more developing countries will take the place of the developed ones. It is natural that Africa attracts more attention, and we have more developing countries in the recommend list.



## 6. Further Discussion

Because the Location Determination Model produces just passable result through the solution oriented by language structure, we apply the latter solution oriented by geographic condition to further discuss whether 5 or even less offices to open can be applicable.

Since this solution does not make any restriction on the number of new office, the company can choose to open any arbitrary number of office based on its actual financial condition. As long as the company has a preliminary planning of how many new offices to open and which continent they are located at, we can provide the exact location of each office immediately.

For example, suggest the company decides to cut an office in Europe and open 5 offices in total, it won't affect the location of other offices in other continents, but may change the location of Europe. Calculating this situation on Matlab program, we get the result of short-term and long-term decision shown in the following figure.

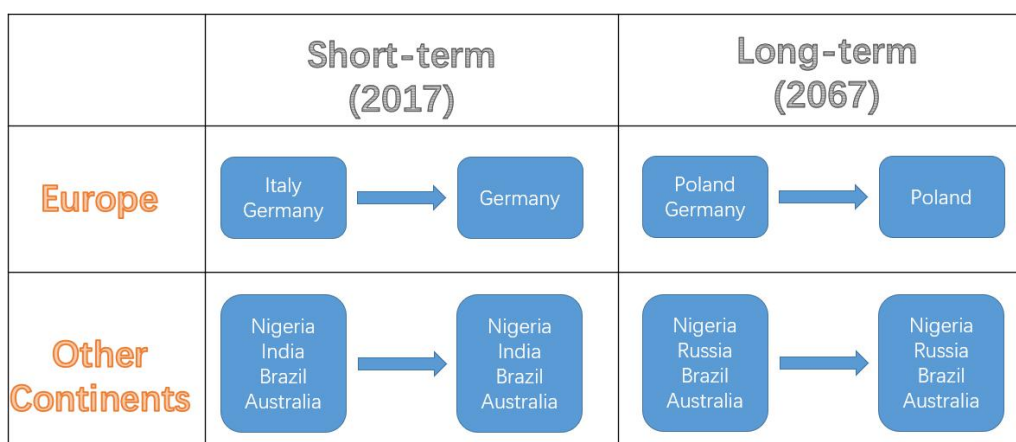


Figure 14 Changes in the location of Europe continent

Therefore, we can say that our model with the geographic-condition-oriented solution has strong adaptability to various situations and is a simple but also feasible method to provide advice on selection of new offices location.

## 7. Strengths and Weaknesses

### 7.1 Strengths

- Based on the AHP method, the Language Development Model fully considers the influence of many possible factors on language development, including communication, geography, economy, academy and diplomacy. Therefore, the result of the model is highly credible.
- The Language Development Model takes country rather than language as the unit of simulation. So, it can not only model the development of a certain language over time, but also model the linguistic structure of any given country over time.
- We looked into the location determination problem from two different aspects: the similarity of linguistic structure and the geographic distance between countries. We made a judgement on the priority of these two factors and improved our model.
- The Location Determination Model can reflect the economic consideration of a modern international company as GDP plays an important role when determining the location for new



offices.

- Flexibility and adaptability are also advantages of the Location Determination Model. When the company change its strategy, new decisions can be quickly derived without modifying the algorithm.

## 7.2 Weaknesses

- In our model, to simplify our model, only the language with the largest number of speakers is considered as the native language of a country. But in fact, there might be two or even more native languages in a country.
- The objective function to optimize the position of new offices is determined with a certain degree of arbitrariness. It can be further improved to meet different actual needs.
- The correctness of our conclusions relies heavily on the accuracy of the data collected and the estimates of GDP, language and population development.
- The recommendation provided by our model remains at the stage of countries (capitals). Further analyses are required in order to determine the specific cities.

## 7. Conclusion

To predict the future distribution of commonly used languages, we build a Language Development Model, setting up 2 criteria to estimate the percentage of speaking a second language, 5 factors to evaluate which second language to speak. Then we simulate the model on Python, using data of population, GDP, language structure and etc. of over 50 countries we collected on the Internet. We conclude that only moderate changes occur to top ten language list, and most of the 16 languages we discuss gain a growth in their total speaker number.

To further provide advice to the multinational service company, we build a Location Determination Model, applying two solutions to figure out the suitable locations for establishment of new offices. The first solution oriented by language structure produced just passable result without comprehensive consideration on economic condition. The second solution, if provided initial number of office on each continent, produces generally ideal locations, and new decisions can be quickly derived if the company change its strategy.

## Reference

- [1] <http://data.un.org/Data.aspx>
- [2] <https://www.cia.gov/library/publications/resources>
- [3] [https://w3techs.com/technologies/overview/content\\_language/all](https://w3techs.com/technologies/overview/content_language/all)
- [4] <http://akarlin.com/2011/05/top-10-most-useful-languages/>
- [5] Kai L. Chan, PhD, POWER LANGUAGE INDEX, May 2016
- [6] Gustafson, D. E. (1979). Fuzzy clustering with a fuzzy covariance matrix. Proc of IEEE Cdc San Diego Ca, 2, 761-766.

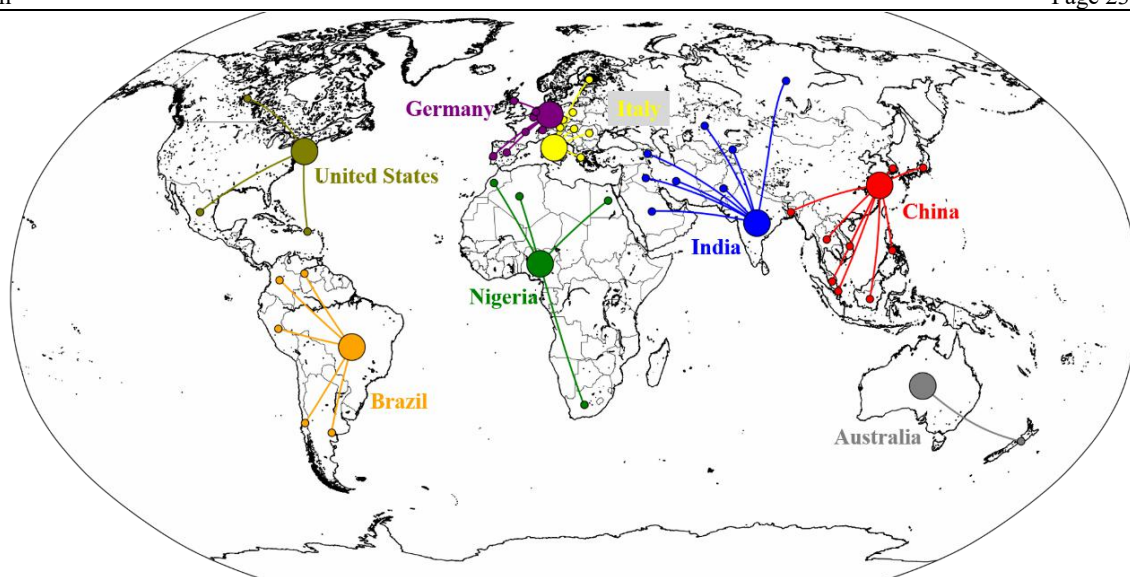
## Memo to the Service Company

### Dear Chief Operating Officer,

We are glad to hear your recent plan about opening six new international offices, which will definitely be an important milestone for your company to become truly international. For such a significant plan, it is necessary to carry out some research before making the final decision. To assist you to decide the location of new offices and determine the languages spoken in the offices, we performed investigations using mathematical modelling.

Firstly, we build up a Language Development model to investigate the distribution of various language speakers over time. In this model, we consider many possible influencing factors, including economy, geography, communication, knowledge and diplomacy. This model is based on data from the United Nation (UN). Through our calculations and tests, we found that this model can really help us to predict the future of language development. It is indicated in this model that over the next 50 years, two original most spoken languages (Russian and Portuguese) will be replaced (by German and Japanese), while the total number of speakers of most languages will gain an impressive increase.

Based on the model of language distribution, combined with Algorithm for Fuzzy Clustering method and P-Center method, we develop a Location Determination Model using two solutions to further provide a recommendation for your company to select the location of your new international office and its working language. One of the solution is oriented by language structure, and the other by geographic condition. Taking language distribution as well as economic development of different countries into account, we suggest to locate new offices in the capitals of Nigeria, India, Germany, Italy, Brazil, Australia for short-term interest. These six locations as well as the languages to be used in the new offices are as the following:



| Recommended Location for New Offices | Language(s) to be Used     |
|--------------------------------------|----------------------------|
| New York, America                    | Spanish                    |
| Shanghai, China                      | Chinese and Japanese       |
| Moscow, Russia                       | Russian and Hindustani     |
| Warsaw, Poland                       | German                     |
| Berlin, Germany                      | German, French and Spanish |
| Abuja, Nigeria                       | Arabic                     |
| Brasilia, Brazil                     | Spanish                    |
| Canberra, Australia                  | Malay                      |

Considering from long-term perspective, your company should build these six new offices in the capitals of Nigeria, Russia, Germany, Poland, Brazil and Australia. Moreover, since our model makes no restriction on number of new office, you can set arbitrary number of offices based on your actual financial condition. As long as you make a preliminary planning of how many new offices to open and which continents they are located at, we can provide the exact location of each office immediately.

We hope that our suggestion can provide useful information for you.

Yours sincerely,  
Team #77238