Statistical Model for the Trend of Prevalent Languages Speakers

Yi Xie, Xiaojie Lei, and Peiai Zhang

***Abstract*—With the advent of increasingly accelerated globalization, the trend of prevalent languages is changing to meet the rapid development. This paper aims to analyze the change of the number of languages speakers considering the effect of the growth of global population and migration in next 50 years. Firstly, we focus on native language speakers and second language speakers in local areas to analyze the number of language speakers in 50 years. We adopt method of social science to detect the main factors: difficulty of language, area where languages are promoted by the government, comprehensive national strength and rates of natural increase. Specific time series forecasting methods are applied to predict the number of language speakers as the last two factors vary through time. Secondly, we introduce migration rate to describe the effect of migration and population in the countries of destinations. Then we model the impact of migration in the second language speakers by selecting the areas that are influenced by migration and sorting these areas into two kinds, Exchange Type and Circle Type, which reflects how migration patterns influence the number of second language speakers in these areas over the next 50 years.**

***Index Terms*—Clustering algorithm, first and second languages, migration patterns, population of language speakers, time series forecasting methods.**

I. Introduction

There are a great number of languages in the world. Native language is a language that a person has been exposed to from birth or within the critical period, and the second language is an auxiliary language for people to acquire [1] and use in the language environment [2]. About half the world's population claim one of the following ten languages (in order of most speakers) as a native language: Mandarin (incl. Standard Chinese), Spanish, English, Hindi, Arabic, Bengali, Portuguese, Russian, Punjabi, and Japanese [3].

It is common that the majority of the world's population also speaks one another language, which is called as second languages. Take an example. No matter where you are, even if you don't speak the official language in local areas, there's a high chance that you are able to communicate with local residents using the second most spoken language [4]. Therefore, we measure the trend of popular languages by mainly considering the number of first and second languages speakers in the paper.

The difference [5] and the contact [6] between the first language and second language have been highlighted by many researches. Language change is accelerated by language contact, especially by contact that occurs when a group of speakers shift from one place to another. This has commonly been explained by linguistic innovation occurring during second language acquisition [7]. This hypothesis is based on historical reconstructions of instances of contact and has not been formally tested on empirical data.

A number of studies about the trend of language changes are based on linguistics and social science. In the paper, we summarize the reasons for the trend of language changes in two dimensions: internal factors and external factors. For one thing, the number of language speakers changes not only because of the natural increase in population [7] but also these new second language speakers in local areas. For another, the number of language speakers, especially the second language speakers, is influenced by international migration and accommodation [7].

We argue that an understanding of the faculty of language requires substantial interdisciplinary cooperation [7]. Some papers apply mathematical models [7], which usually focus on how a specific dialect can spread across an area or how the distribution of a dialect is. For example; Cellular automata were used to describe the use of cellular automata to model dialect feature diffusion as the adaptive aspect of the complex system of speech [7].

This paper ingeniously combines the methods of linguistics, social science, and statistic models to describe the change in the number of language speakers globally. First, we divide the languages speakers to native language speakers and second language speakers and analyze the increase in local areas.

Second, we consider that the international migration also plays a critical role in the growth of the numbers of second languages. Therefore, we define migration rate to select the areas where have great amount of migration.

II. Justifications and Notations

The extraneous factors like government's language policy, requirement for study, employment, the trend of the times and so on, mainly influence the number of second language speakers. Because the affected population have usually mastered a language.

The change of the distribution of languages is mainly produced by population migration. Reasoning that the situation of introducing non-native languages in the region is extremely minor.

Migration model does not consider policy issues, such as the United States banning Mexican immigrants.

Manuscript received January 25, 2019; revised March 18, 2020. This work was supported in part by the Mathematics Department, Jinan University.

The authors are with the Mathematics Department, Jinan University, No.601, West Huangpu Avenue, Guangzhou, Guangdong, China (e-mail: xie@stu2016.jnu.edu.cn, littlegrace163@163.com, tzhangpa@jnu.edu.cn).

doi: 10.18178/ijlll.2020.6.1.245

The rank of the difficulty of language and the area where the governments promote will not change obviously in a short time.

We only consider the 26 languages in the 2017th edition of Ethnologue1. Because these 26 languages have 50 million or more total speakers and have a large impact all over the world.

Symbol	Description
INS	the index of comprehensive national strength
RDL	the rank of difficulty of language
GPA	the government promotion area
CD	countries of destinations
CO	countries of origins
EPCO	the emigrant population from countries of origins
RNGCD	the rate of natural increase of countries of destinations
RNGCO	the rate of natural increase of countries of origins
MR	the migrant rate
PNL	the future population of native language speakers
PSL	the future population of second language speakers
TLS	the total population of language speakers

III. Predictive Model 1: The Local Number of Language Speakers

Firstly, we predict the local number of language speakers divided into two groups, the native language speakers and the second language speakers. The number of native language speakers in local areas is mainly influenced by the natural growth rate. As for the number of second language speakers, many factors have great impact on it. According to the social science, we detect three main factors, the index of comprehensive national strength, the rank of difficulty of languages and government promotion.

A. The Number of Native Language Speakers in Local Areas

We analyze the natural population growth of the major countries where a certain language is the native language. We can get the natural growth rate of the native language speakers in the future through data fitting, and predict the future population of the native language speaker.

$$\text{Growth Rates} = \frac{\text{This Year's Population - Previous Year's Population}}{\text{Previous Year's Population}} \quad (1)$$

We finally get the number of native language speakers through iteration and the rank in next 50 years (see Table I).

TABLE I: Prediction of the Top-Ten Number of the Native Language Speakers

Language	2017 YEAR	2017 RANK	2027 RANK	2047 RANK	2067 RANK	2067 YEAR
Arabic	290000000	5	2	1	1	1947210800
Hindustani (Hindi/Urdu)	329000000	4	3	3	2	1288489020
Bengali	242000000	6	7	2	3	961693480
Mandarin Chinese	897000000	1	1	5	4	922459300
Portuguese	218000000	7	5	4	5	812190703
English	371000000	3	6	7	6	712810462
Spanish	436000000	2	4	6	7	686833416
Hausa	85000000	11	9	8	8	678218400
Punjabi	148000000	9	8	9	9	579624240
Malay	77000000	15	10	10	10	459482100

B. The Number of Second Language Speakers in Local Areas

1) Qualitative analysis

In the research, the following three indicators are considered as influential factors of change in the number of second language speakers: the index of comprehensive national strength, the rank of difficulty of languages and government promotion.

a) The Index of Comprehensive National Strength (INS)

The spread and development of second language are strongly relevant to the strength of the country [7]. The language of economically developed countries will spread more widely for economic, social, political and other reasons. INS of the country obviously changes in 50 years. We predict INS by different time series analysis according to the trait of the past years. The data provided by Frederick S. Pardee Center2 for International Futures are referred to quantify INS.

b) The Rank of Difficulty of Languages (RDL)

In general, the difficulty of the language itself affects the number of second language speakers. People tend to learn a relatively simple language. The harder it is, the less likely people in other countries are willing to learn. WeirdIndex3, the indicator derived from The World Atlas of Language Structures (WALS), is used to evaluate the difficulty of language. The higher WeirdIndex value is, the weirder the language is, that is, relatively the higher difficulty the language is.

c) The Government Promotion Area (GPA)

To a certain extent, the degree to which the government promotes national language also affects the spread of it. When the government vigorously promotes national language, it will arouse the enthusiasm of people in other countries to learn it as a second language, so that the number of this language speakers will increase. The total areas4 (except its own country) where this language is promoted are used to quantify to which extent the government promotes their national language.

2) Time series analysis

We use time series analysis method to predict INS in the country. Because of the different development situations in different countries, time series shows different trends. Therefore, we apply different models for actual situations, respectively simple moving average method, trend moving average method and exponential smoothing method [7] .

I. When the comprehensive national power index, due to slow development, doesn't show obvious trend, we use moving average method to eliminate the impact of uncertain factors, analyzing and predicting long-term trends in the sequence. In the situation that time series has no obvious trend changes, simple moving average method can accurately reflect the actual situation. Note sequence $y_1, y_2, \ldots, y_T$. Use these data to predict the next data y_{T+1}.

i) According to the calculation formula of once simple moving average, we predict the forecasting sequence. N is a parameter which is a positive integer. N∈ [5,200] and N<T.

$$\hat{y}_{t+1}(N) = \frac{1}{N}(y_t + y_{t-1} + \ldots + y_{t-N+1}) \tag{2}$$

Thus, we get the predicted sequence, $\hat{y}_{N+1}(N), \hat{y}_{N+2}(N), \ldots, \hat{y}_T(N), \hat{y}_{T+1}(N)$.

ii) Accurate the standard deviation to decide the parameter N. The formula of the standard deviation depends on the only parameter N.

$$S^{(1)}(N) = \sqrt{\frac{\sum_{t=N+1}^{T}(\hat{y}_t(N) - y_t)^2}{T-N}} \tag{3}$$
$$N^* = \min \arg\{S(N) \mid N = 5, 6, \ldots, 9, 10\}$$

iii) After we find out the most accurate parameter N^*, we can calculate the $\hat{y}_{T+1}(N^*)$. We repeat the method mentioned above and calculate the forecasting sequence until the M year we want.

$$\hat{y}_{T+1}, \hat{y}_{T+2}, \ldots, \hat{y}_M$$

II. However, when the time series shows a trend of linear increase or decrease, hysteresis deviation can be predicted by using trend moving average method. The theory of the trend moving average method is on the basis of moving average method.

i) Predict the first forecasting sequence. In this part, as long as we simply repeat the moving average method, we get the first forecasting sequence.

$$\hat{y}^{(1)}_{t+1}(N) = \frac{1}{N}(y_t + y_{t-1} + \ldots + y_{t-N+1}) \tag{4}$$
$$t = N, N+1, \ldots, T$$

$$S^{(1)}(N) = \sqrt{\frac{\sum_{t=N+1}^{T}(\hat{y}^{(1)}_t(N) - y\_t)^2}{T-N}} \tag{5}$$
$$t = N, N+1, \ldots, T$$
$$N^* = \min \arg\{S^{(1)}(N) \mid N = 5, 6, \ldots, 9, 10\}$$

$$\hat{y}^{(1)}_{N+1}, \hat{y}^{(1)}_{N+2}, \ldots, \hat{y}^{(1)}_T, \hat{y}^{(1)}_{T+1}, \ldots, \hat{y}^{(1)}_M$$

ii) Predict the second forecasting sequence. We just regard the first forecasting sequence as the original data. Use the data to predict the next forecasting sequence. We can simply use the same parameter N^*.

$$\hat{y}^{(2)}_{t+1}(N) = \frac{1}{N}(\hat{y}^{(1)}_t + \hat{y}^{(1)}_{t-1} + \ldots + \hat{y}^{(1)}_{t-N+1}) \tag{6}$$
$$t = 2N^*, 2N^* + 1, \ldots, T$$

$$\hat{y}^{(2)}_{2N^*+1}, \hat{y}^{(2)}_{2N^*+2}, \ldots, \hat{y}^{(2)}_T, \hat{y}^{(2)}_{T+1}, \ldots, \hat{y}^{(2)}_M$$

iii) According to the formula of the trend moving average method:

$$\hat{y}_{T+m} = a_T + b_T\, m, m = 1, 2, \ldots, M - T$$
$$a_T = 2\hat{y}^{(1)}_T - \hat{y}^{(2)}_T \tag{7}$$
$$b_T = \frac{2}{N^*-1}\hat{y}^{(1)}_T - \frac{2}{N^*-1}\hat{y}^{(2)}_T$$

III. When the development of the country is rapid, exponential smoothing method can satisfy the trend and describe it better.

i) In exponential smoothing method, the next data is directly related to the previous one. The formula is

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t, \alpha \in [0,1] \tag{8}$$

α is a parameter like N mentioned above.

ii) Decide the initial term y^_1. The initial term y^_1 depends on the original time sequence. You can define it as $\hat{y}_1 = y_1$ or $\hat{y}_1 = \frac{y_1+y_2}{2}$.

iii) Accurate the standard deviation to decide the parameter α. The formula of the standard deviation depends on the only parameter α.

$$S(\alpha) = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t(\alpha) - y_t)^2}{T}} \tag{9}$$
$$\alpha^* = \min \arg\{S(\alpha) \mid \alpha = 0.1, 0.2, \ldots\}$$

After we determine the value of α, we can get the forecasting sequence.

$$\hat{y}_{T+1}, \hat{y}_{T+2}, \ldots, \hat{y}_M$$

Besides, we also can predict the second forecasting sequence based on the first forecasting sequence.

3) Quantitative model

Through the data obtained before, we establish the relation between three factors and the number of second language speakers through curve fitting. RDL and GPA will not change obviously in a short time. We conclude that changes of INS contribute to main changes of the number of second language speakers.

Using the toolbox of MATLAB we make data fitting and gain the relation between three factors and the number of second language speakers. The principle is to use Delaunay triangles for spatial reconstruction and do interpolation operation on the surfaces of the triangle.

The Delaunay triangle consists of a series of continuous polygons that consist of vertical bisectors. N points that differ on a plane, divided the plane by the nearest principle, and each point is associated with its nearest neighbor [7].

In the fitting model, we regard RDL, GPA and INS as independent variables. PSL is considered as a dependent variable. We observe the trend of the data, then fit it appropriately.

Considering the change in the population of second language speakers, we apply the exponential smoothing method and the choice of weighting coefficient is particularly important. The value of α illustrates the proportion of deviation in the new forecast. The larger α is, the smaller the

proportion of the original predicted value is. If equation (5) rewritten as

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t) \quad (10)$$

It can be seen that the new predicted value is obtained by correcting the original predicted value according to the prediction error. The magnitude of α reflects the magnitude of the correction. Therefore, the larger the α, the larger the magnitude of the correction.

We finally obtain the number of the second language speakers and the rank in 50 years (see Table II).

TABLE II: PREDICTION OF THE TOP-TEN NUMBER OF THE SECOND LANGUAGE SPEAKERS

Language	2017 YEAR	2017 RANK	2027 RANK	2047 RANK	2067 RANK	2067 YEAR
English	611000000	1	1	1	1	2306624336
Arabic	132000000	6	4	2	2	323380733.6
Malay	204000000	3	2	3	3	243759286.9
Russian	113000000	7	3	4	4	178985539.4
Mandarin Chinese	193000000	4	5	9	5	170349471.8
Hindustani (Hindi/Urdu)	215000000	2	6	6	6	120000000
Spanish	91000000	8	11	5	7	92604534.86
Swahili	91000000	9	7	8	8	91000000
French	153000000	5	8	7	9	88465544
Hausa	65000000	10	9	10	10	65000000

IV. PREDICTIVE MODEL 2: THE GLOBAL NUMBER OF LANGUAGE SPEAKERS

Model 1 reveals that INS, RDL and GPA are the intrinsic factors in terms of the number of languages speakers while Model 2 considers the extrinsic factors. The number of language speakers are also influenced by the international migration from the outside. This model takes the migration into account. Global population and human migration accelerate the spread of languages and enhance the number of the second language speakers in other places by bringing new languages to these areas but the number of native language speakers are seldom influenced [7]. Immigrants pass their languages on to other members of their speech community.

A. Migration in the Second Language Speakers

We analyze migration by destinations and origins, finding out the trends and patterns [7] in international migrants. Countries of destinations (CD) whose migrant stocks are relatively huge are chosen in the past. In other words, areas whose scale of emigrant population is so small that they are not affected by migration patterns. Therefore, we use clustering algorithm to find out the most influential countries of origins (CO). Clustering algorithm [7] sorts the migrant stocks data into two kinds, effective or ineffective.

We can analyze the change of second language speakers worldwide and just take Russian for example. Then, e calculate and predict the number of immigrants5 coming from other CO to Russia from 1900 to 2015 (see Fig. 1. The x axis represents countries of origins. Every country is sorted according to country code, ranging from 1 to 250). For Russian, the most influential countries of origins are Kazakhstan (No.110) and Ukraine (No.218).

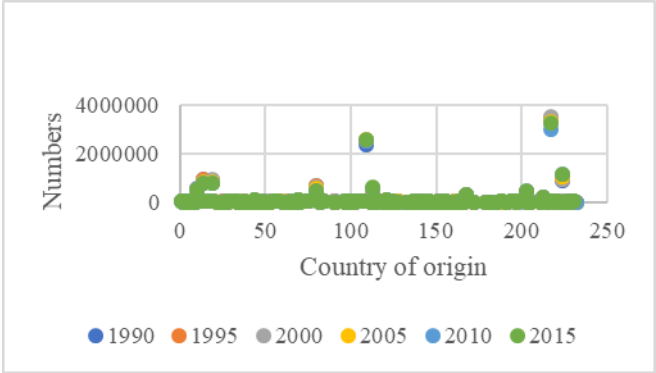

Fig. 1. Migrant stock of Russian federation.

The emigrant population from countries of origins (EPCO) plays an important role in the spread of new languages in local areas. The population of CD and CO makes contributions to the importance of emigrant population, so the future population of native language speakers (PNL) in destinations and origins should be involved. The rate of natural increase of countries of destinations (RNGCD) and the rate of natural increase of countries of origins (RNGCO) measure PNL in destinations and origins. The migrant rate (MR) is defined to analyze the influence.

$$MR_t^{\text{destinations and origins}} = \frac{EPCO_t \times RNGCO_t}{PNL_t^{\text{destination}} \wedge NGCD_t}, t = 1990, 1991, \ldots, 2050 \quad (11)$$

If MR is huge for a long time, the impact of the language from CO should be influential in CD, which results to the boost of the number of corresponding second language speakers. Time series forecasting is used to predict MR. For example, we use moving average method to eliminate the impact of uncertain factors, analyzing and predicting long-term trends in $MR_t^{Russia\_Kazakhstan}$ and $MR_t^{Russia\_Ukraine}$, which are high and grow slowly (see Fig. 2). It reflects the trend of the immigration from Kazakhstan and Ukraine can enhance the influence of their native languages in Russia. In other words, the numbers of people who speak Kazakh and Ukrainian in Russia increases [7].

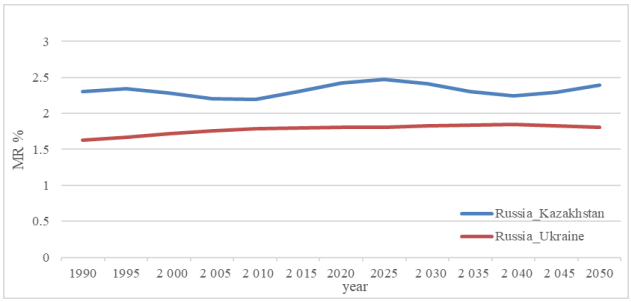

Fig. 2. $MR_t^{Russia\_Kazakhstan}$ and $MR_t^{Russia\_Ukraine}$.

V. RESULTS

A. Results of the Number of Language Speakers in Local Areas

After obtaining the number of the native and second

language speakers, we calculate the total number of the language speakers in local areas. At the same time, we rank it as the top-ten most language speakers in total (see Table III).

$$TLS=PNL+PSL \tag{12}$$

TABLE III: COMPARISON OF THE TOP-TEN NUMBER OF LANGUAGE SPEAKERS

Language	2017 YEAR	2017 RANK	2067 RANK	2067 YEAR
English	983000000	2	1	3019434798
Arabic	422000000	5	2	2270591534
Hindustani (Hindi/Urdu)	544000000	3	3	1408489020
Mandarin Chinese	1090000000	1	4	1092808772
Bengali	261000000	8	5	972566904
Portuguese	229000000	9	6	819141479
Spanish	527000000	4	7	779437951
Hausa	150000000	11	8	743218400
Malay	281000000	6	9	703241387
Punjabi	148000000	12	10	585705030

Through the Table III, the population of various language speakers has apparently changed after 50 years.

First, there are 8 languages still in the top ten lists of 2067. For the reason that China's population growth has been well controlled, the number of Chinese speakers has relatively little growth so the rank will decline. The rank of English and Hindustani will remain steady. Second, since Hausa and Punjabi population growth in the region keeps high over time, it will be in the top-ten lists in 50 years. In addition, there is a sharp increase of the Arabic speakers in recent years. The deviation exists when we make a curving fitting but it still proves a boom of Arabic.

Third, it is expected that more proportion of world's population will claim the top 10 prevalent languages. This trend reflects the fact that descendants of people who speaker minority languages might shift their first languages from the minority to the majority.

B. Results of the Global Number of Language Speakers

The original distribution of languages in 2017 is shown (see Fig. 3).

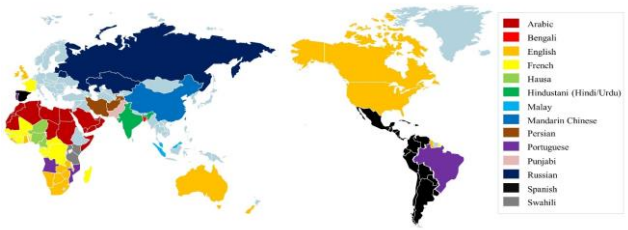

Fig. 3. Language distribution in 2017.

We sort the countries of origins which are impacted by the migration into two categories, Exchange Type and Circle Type (see Table IV). Exchange Type represents that the countries of destiny and origin exchange their languages. The Circle Type represents that the country of destiny only absorbs new languages from country of origins. We use the map to describe the different types of areas around the world generally (see Fig. 4).

TABLE IV: TWO TYPES OF THE AREAS (INFLUENCED BY THE MIGRATION)

Type	Areas
Exchange Type	Argentina
	Australia
	Belarus
	Britain
	France
	Greece
	Italy
	Kazakhstan
	Malaysia
	New Zealand
	Poland
	Portugal
	Russia
	Singapore
	Spain
	Switzerland
	Ukraine
	Uzbekistan
	Venezuela
Circle Type	America

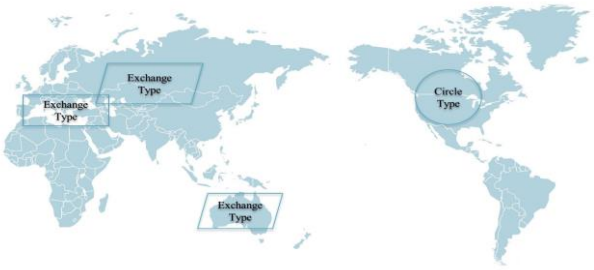

Fig. 4. The Category of the area.

Take Russia (Exchange Type) and America (Circle Type) for example.

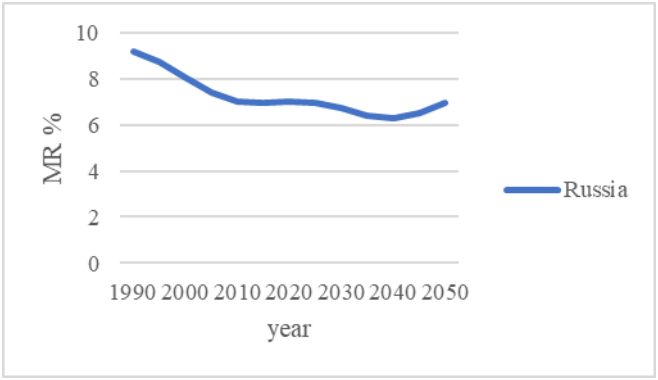

Fig. 5. $MR_t^{Ukraine\_Russia}$.

As for Exchange Type areas like Russia, the immigrants bring the new languages to this place, which enhances the number of second language speakers in these areas. As we have mentioned above, the migration from Ukraine and Kazakhstan plays a critical role in the number of second

language speakers in Russia. The migration from Russia also influences the number of second language speakers in Ukraine because $MR_t^{Ukraine\_Russia}$ is quite high (see Fig. 5). The change of Russian is accelerated by language contact, especially by contact that occurs when a group of speakers acquire Ukraine. This has commonly been explained by linguistic innovation occurring during second language acquisition [7]. Besides, Russian also has great effect on Ukraine though this trend might decline.

What's more, Ukraine gained its independence from the Soviet Union in the aftermath of its dissolution at the end of the Cold War. Ukraine has maintained its independence as ever since [7]. The political relationship between the Ukraine and Russia is complex, which makes contributions to the communication between the two languages in a degree. According to the migrant rate (see Fig. 2), we can draw the conclusion that the Kazakh and Ukrainian expand in Russia though the trend is raising slowly.

America has diverse languages, resulting the fact that the mainstream of languages is multilingual. Mexico has the largest emigrant population to America, the Circle Type, keeping growing in recent years. However, the number of second language speakers in America is too diversified to be influenced by Spanish. Therefore, although the migration from Mexico to America is large, the numbers of every languages speakers in America keeps its diversity and merely change (see Fig. 6). What's more, as the new immigration policy changes [7], the emigrant population might decline in the future.

Language contact occurs when speakers of two languages interact but leads to different results. The languages in Russia and the languages in Ukraine affect each other. America has diversity of languages, resulting the fact that the mainstream of languages is multilingual.

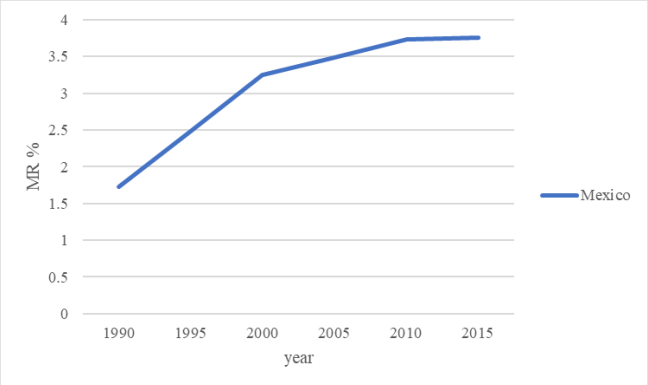

Fig. 6. $MR_t^{Mexico\_America}$.

VI. Conclusion

In the paper, firstly we develop the prediction model about the number of language speakers in local areas with qualitative and quantitative analysis. Based on our forecast, there will be noticeable changes in the numbers of native speakers and second language speakers in the next 50 years. The rank of English and Hindustani will remain steady while Mandarin Chinese will decline. A future top-ten list shows that French and Russian in the current top-ten lists will be replaced by Hausa and Punjabi due to the high growth of population in these two areas. Then we take migration into count to study the geographical language distribution. We sort these areas influenced by migration into Exchange Type and Circle Type. The former refers to areas that have great migration and influence the countries of origins. The latter refers to areas which have massive migration and are influenced by the new language from the countries of origins. The migration from China [7] and Malaysia is predicted to make significant impact, resulting in the increase of the second language (Mandarin Chinese and Malay) speakers in the South Asia. The emigration from Russia will make contributions to the expand of corresponding second language speakers around the border of Russia. As for America, because of its diverse languages, the mainstream of languages is still the same and not impacted by the languages from the countries of origins. In brief, the prediction analysis shows the trend of language development in the future.

Notes

A. We refer to the magazine, Ethnologue, which is edited by Lewis Paul M. (16th ed; Publisher: SIL International, Dallas, 2009.)
B. The raw data of power measure are retrieved from the public domain webpage http://www.ifs.du.edu/ifs/frm_PackagedDisplaysFlex.aspx.
C. The raw data of WeirdIndex is retrieved from the public domain webpage https://wals.info.
D. The raw data of the total areas is retrieved from the public domain webpage https://www.wikipedia.org.
E. The raw data is retrieved from the public domain webpage http://www.un.org/en/development/desa/population/migration/data.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Peiai Zhang conducted the research; Yi Xie analyzed the data; Xiaojie Lei wrote the paper; all authors had approved the final version.

References

[1] L. F. Diane, "Looking ahead: Future directions in, and future research into, second language acquisition," *Foreign Language Annals*, vol. 51, no. 1, pp. 55-72, 2018.
[2] S. L. Thorne, R. W. Black, and J. M. Sykes, "Second language use, socialization, and learning in internet interest communities and online gaming," *Modern Language Journal*, vol. 93, pp. 802-821, 2009.
[3] J. Lane, "The 10 most spoken languages in the world," *Report of Babbel Magazine*, Nov. 14, 2018.
[4] K. A. Noels, L. G. Pelletier, R. Clement, and R. J. Vallerand, "Why are you learning a second language? Motivational orientations and self-determination theory," *Language Learning*, vol. 53, pp. 33-63, 2003.
[5] J. S. Johnson and E. L. Newport, "Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language," *Cognitive Psychology*, vol. 21, pp. 60-99, 1989.
[6] M. S. Schmid and B. Kopke, "The relevance of first language attrition to theories of bilingual development," *Linguistic Approaches to Bilingualism*, vol. 7, no. 6, pp. 637-667, 2017.

[7] C. Yang, "A formalist perspective on language acquisition," *Linguistic Approaches to Bilingualism*, vol. 8, no. 6, pp. 665-706, 2018.
[8] C. J. Bajema, "Estimation of the direction and intensity of natural selection in relation to human intelligence by means of the intrinsic rate of natural increase," *Eugenics Quarterly*, vol. 10, pp. 175-187, 1963.
[9] L. B. Anderson, *Migration, Accommodation and Language Change*, 1st ed. London: Palgrave Macmillan, 2008, pp. 179-183.
[10] M. D. Hauser, N. Chomsky, and W. T. Fitch, "The faculty of language: What is it, who has it, and how did it evolve?" *Science*, vol. 298, no. 5598, pp. 1569-1579, 2002.
[11] G. J. Baxter, R. A. Blythe, W. Croft, and A. J. McKane, "Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English," *Language Variation and Change*, vol. 21, pp. 257-296, 2009.
[12] W. A. Kretzschmar and I. Juuso, "Cellular automata for modeling language change," in *Proc. 11th International Conference on Cellular Automata for Research and Industry*, vol. 8751, 2014, pp. 339-348.
[13] N. V. Borisova, "The strength of territorial autonomy and language preferential policy in contemporary world," *Sravnitelnaya Politika-Comparative Politics*, vol. 9, no. 2, pp. 62-82, 2018.
[14] B. D. Fulcher, M. A. Little, and N. S. Jones, "Highly comparative time-series analysis: The empirical structure of time series and their methods," *Journal of the Royal Society Interface*, vol. 10, no. 83, 2013.
[15] S.-W. Cheng and T. K. Dey, "Quality meshing with weighted Delaunay refinement," *Society for Industrial and Applied Mathematics*, vol. 33, no. 6, pp. 69-93, 2003.
[16] D. Lipowska and A. Lipowski, "Language competition in a population of migrating agents," *Physical Review E.*, vol. 95, no. 5, 2017.
[17] G. J. Abel and N. Sander, "Quantifying global international migration flows," *Science*, vol. 343, no. 6178, pp. 1520-1522, 2014.
[18] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3336-3341, 2009.
[19] D. Arel, "Demography and politics in the first post-Soviet censuses: Mistrusted state, contested identities," *Population*, vol. 57, no. 6, p. 791, 2002.
[20] T. Leuschner, "Sarah G. Thomason, language contact: An introduction (Edinburgh: Edinburgh UP, 2001)," *Linguistik Online*, 2001.
[21] D. H. Struk, *Encyclopedia of Ukraine*, 2nd ed. University of Toronto Press, 1993, ch. 5.
[22] J. A. Garcia, "Political integration of Mexican immigrants: Explorations into the naturalization process," *The International Migration Review*, vol. 15, no. 4, pp. 608-625, 1981.
[23] R. J. LaPolla, "Language contact and language change in the history of the Sinitic languages," *Harmony of Civilization and Prosperity for All*, vol. 2, no. 5, pp. 6858-6868, 2010.

Yi Xie was born in Zhejiang, China in 1997. She is an undergraduate student at the Mathematics Department, Jinan University in Guangzhou, China. She has presided over or participated in 2 research projects at both university and national level. Her current research interests cover machine learning and statistics.

Xiaojie Lei was born in Guangxi, China in 1997. She is an undergraduate student at the Mathematics Department, Jinan University, in Guangzhou, China. She has presided over or participated in 2 research projects at both university and national level. Her current research interests cover applied statistics.

Peiai Zhang was born in Shandong, China in 1973. She got her bachelor degree of mathematics in 1996 at Shandong University in Jinan, got her master degree of computation mathematics in 1999 at Dalian University of Techlogy in Dalian, and gained her doctor degree of operations research and control theory in 2002 at Dalian University of Techlogy in Dalian. She has been working in Jinan University from 2002. She is interested in optimization as an associate professor.