Summary

This report presents a model of the distribution of the worlds most important languages over the next 50 years. With insight and data from the past 50 years, we built a model that captures how countries develop over time, affected by both internal factors and influence from other countries. With particular interest in existing linguistic diversity, immigration patterns, business relations, media production and consumption, and use of technology, we maximize the usefulness of available data and produce results that align closely with existing research and predictions. Using our predictions, we advise on regions of interest for a growing international company.

To accomplish our goals, we first researched the factors that heavily influence the spread of a language. We boiled them down to the indicators listed above, and then weighed them against each other to produce an influence metric between every country of interest. We then paired all these influence metrics with each country's individual growth patterns to output a language profile for every country in the year 2068. Not only does this approach give us a prediction for the entire world, but it also allows us to key in on specific regions and trace their influences by looking at which factors affected each country the most.

Our results principally showed an unsurprising increase in English and Mandarin. English develops all over the world, while Mandarin's growth is more localized to China, with reasonable growth in the many countries China does business with. We don't see a single large increase in Arabic, but rather a moderate increase in almost every country. Our model does not show Spanish becoming as prolific as some researchers have suggested, as the majority of Spanish speaking countries are not growing as quickly as countries in the developing world, particularly India (Hindi continues to outpace Spanish). Accordingly, we see a massive increase in English in Western Africa, and an increase in Bengali, Malay, Punjabi, and Portuguese. We see Russian decline and fall out of the top ten languages spoken worldwide, while English soars over Mandarin, likely due to to its status as the global lingua franca.

The following paper details our choice of metrics and mathematical modeling, our most interesting results, and areas in which we feel our model is particularly strong or could be improved.

Memo

Dear Chief Operating Officer,

The following is a recommendation for the locations to open a new headquarters in and what languages to tailor the hiring of our new employees toward. This is based on a projection of what the international market will look like over the next 50 years. Projecting these trends should aid in our understanding of how to maximize long term profit for our company.

Our analysis is based primarily on the long term investment into a region's economy. As a result, our recommendations are influenced accordingly - possibly sacrificing short term growth in preference for long term reward and stability. Influencing our model are primarily the effects of immigration, trade, media consumption, and tech. The majority of which are based on one country's influence on another. Creating an expected growth in a country's population to speak a language because of the influence from another country it interacts with based on these metrics. After simulating these connections for 50 years, a prediction for the growth of languages in countries and regions over that time period can be made.

One important caveat to note in our simulation is differing dialects. For the purposes of projecting language growth in generality, different dialects of languages are ignored. However, this will prove problematic in our company's penetration of those markets, so our recommendations will take that into account and will not be based entirely on the overall growth of a language.

Recommendation 1: Middle East

We project rapid growth in the economies of many Middle Eastern countries, and correspondingly, an international growth of the Arabic language. Arabic is the most popular language in this region and so we recommend it as a language to have employees learn, in addition to English. We do not recommend more than these two languages since Arabic has a few different dialects that it would be important for employees to learn. This would allow a more engrossing market penetration for the region, by allowing us to communicate with the region as a whole. Additionally, we would recommend Egypt to be the location for these new offices, as it contains the largest population of Arabic speakers and is also centrally located for the MENA population in which Arabic is a dominant language. Also, this office could target the English speaking population, centralized in Nigeria, since English is projected to have overwhelming control of this region.

Our reasoning for this recommendation first is that, while Arabic was not the most popular language based on population totals, it grew in almost every demographic across the world. Starting a new office in its centralized area, strategically, makes sense for future penetration into other international markets.

Recommendation 2: Indonesia

Along with our thoughts on the spread of Arabic worldwide, a market that could be accessible because of Arabic, but grown with other languages is Indonesia. While having a modest Arabic population, Indonesia has incredibly large Malay and Mandarin populations that

would help create a concrete target demographic. Geographically, this location is also ideal as it bridges a connection with the Shanghai and Egypt offices.

Recommendation 3: India

Because of its massive population and influence on both neighboring and international markets, India provides another possible location to target. With high populations of Hindi, English, and Bengali speakers, not only would this location allow us to take advantage of the rapid growth of the Indian economy and population, it would also provide a great way to gain a market share in large neighboring countries such as Bangladesh with a high Bengali population. Additionally, it is easy to note that India's rapid population growth and growing tech market could prove to be more influential that our current model predicts.

Recommendation 4: Germany

Germany's economic dominance in Europe proved to be a strong factor in our model, causing many neighboring countries to adopt it as either a primary or secondary language. With Germany's current success and stability in the European economy, it provides a moderately low risk-high reward location for a new office. Subsequently, we would recommend employees of this office location to learn German, English, French, and Japanese. We predict this office to be a great location for market penetration in Europe, with potential to spread to other major global markets.

Recommendation 5: Brazil

An office in Brazil would provide a great location to target the South American market. Employees here speaking English, Portuguese, and Spanish would be able to not only take a large portion of the Brazilian market (almost exclusively Portuguese speakers), but would also be able to target the neighboring countries such as Colombia and Peru which are predominantly Spanish speaking. Additionally, learning Spanish from Portuguese is a relatively easy progression for employees to make due to their historical similarities.

Recommendation 6: Nigeria

Based on our previous prediction of the Egyptian office being able to target this region, this location may not be necessary. However, given more information on the specific influence metrics between these two regions, it may be necessary to target this market with a standalone office. This office would mostly be responsible for learning English - the dominant language in Nigeria based on our model - however, Swahili and French may be additional languages to target. While the growth of Swahili is largely unpredictable due to its nature being primarily a secondary language, French proved to be largely in control of the Democratic Republic of the Congo and subsequently could influence neighboring countries in the area to adopt French as a business language. This office has the potential for large growth over a long term business strategy, as Africa has a largely untapped population, although this would be dependent on the expansion of the African economy.

Abstract

In this paper we detail a model that predicts the languages that will be most spoken and where in the year 2068. It details our research, the decisions we made, and the effectiveness of our final model. We pay particular attention to the 23 most spoken languages and the world's 34 most populous countries. Doing so gives far reaching lingual, geographic, and economic diversity. Our model produces results that are, for the most part, in line with current predictions and research, with a few interesting surprises, as well as regional detail.

1 Introduction

Our main goal is to build an algorithm that models the changes in language demographics across the globe over the next 50 years. Specifically, we are interested in capturing the influence countries have on each other via immigration, trade, and exported media to develop an understanding of how that influence affects language distribution. For example, over the past 50 years, the percentage of English speakers in China and India has increased dramatically in tandem with a surge of trade relations with the United States. With our model, we intend to simulate the next 50 years to predict similar trends, which will allow us to recommend the locations and languages of interest for an international company's new offices.

We decided to measure global trends by focusing on individual countries, taking into account their individual, internal growth and trends, as well as the effect other countries had on them. The model uses data on current language diversity, languages taught in schools, and demographic trends to build a country's profile. Then it evaluates how much every other country influences the original country using data on immigration trends, trade of goods and services, economic growth, technology use, and media consumption.

Rather than simulating every country and all 6000 languages, we limited our model to 34 countries and the 23 most popular languages. Because we were largely interested in what the ten most popular languages are 50 years from now and where the 6 best locations for new offices are, we determined that it would be highly unlikely for countries and languages outside of our dataset to have a large impact. Furthermore, our country selections ensured that we accounted for the majority of people speaking the most popular languages. Countries were excluded when a larger country with more data available could reasonably represent the eliminated country, such as Saudi Arabia being similar enough to the UAE to represent it and several other nations on the Arabian Peninsula.

Something about our progress and results

This paper discusses the development of our model, its technical details, our results and predictions, sources of error, and places it could be improved.

2 Assumptions

1. Due to the massive number of languages found throughout the world and the massive amount of detailed data required to map the interactions of all 193 countries with each other country, we cut our data set and simulation size down to 34 countries and 23 languages. Languages with significant populations that we excluded either had very little data or were determined to be dying out or primarily secondary languages, like Swahili and Wu Chinese. We believe that their exclusion does not affect the final results of the simulation. Once we selected our languages of interest, we gathered data on the 23 most populous countries and looked at how well the top languages were represented. In some cases, we were missing a large percentage of the language's speakers, as they were found in numerous smaller countries. To fill these holes in the data, we selectively sought out countries that had a high number of speakers of the missing language and were highly representative of the countries around them. For example, we had to make adjustments to include more Arabic and French speaking countries, thus Algeria made it onto our list. We ended up accounting for all languages we think have a chance of making it into the top 10 and are able to represent and generalize over most of the world with the countries included.

2. Our simulation makes no attempt to account for war or natural disasters. Doing so with accuracy is essentially unfeasible (and would hold far more value than our model), and randomly generating events would provide inconsistent data. Predictable events we considered but could not commit to without far more research included the opening of North Korea, further regime change in the Middle East and northern Africa, Russian encroachment on Eastern Europe, and potential mass exoduses of refugees to developed countries.

3. There are innumerable factors that contribute to the growth of a language in a certain region, and every country has its own special relationship with foreign cultures. We picked factors that had consistent, available data and were relevant to all countries. There were several factors we could not account for, such as xenophobia and national identity, cultural integration, and difficulty of learning a language. Ultimately, our goal was to capture large-scale global trends, and the factors we chose reflect that. We give a detailed explanation for our metrics in section 4.2.

3 The Model

3.1 Overview

Our Approach We saw fit to measure the change in world language distributions by breaking a global system down into manageable parts: gathering data and analyzing it country by country. The first step we took was to consider the current day growth of languages within a country in isolation. We then ventured to capture the influence of every other country on that original country and project the country's influence as an influence of it's dominant language(s) on the current language populations and growth rates in the original country. This

model format allows us to focus on smaller, significant interactions between countries, versus looking at global trends as a whole. Since every country operates differently, has populations growing at different rates and interacts with other countries and the global population differently, we believe it is more accurate to predict the growth of languages at this level. Our model is outlined in figure 1.

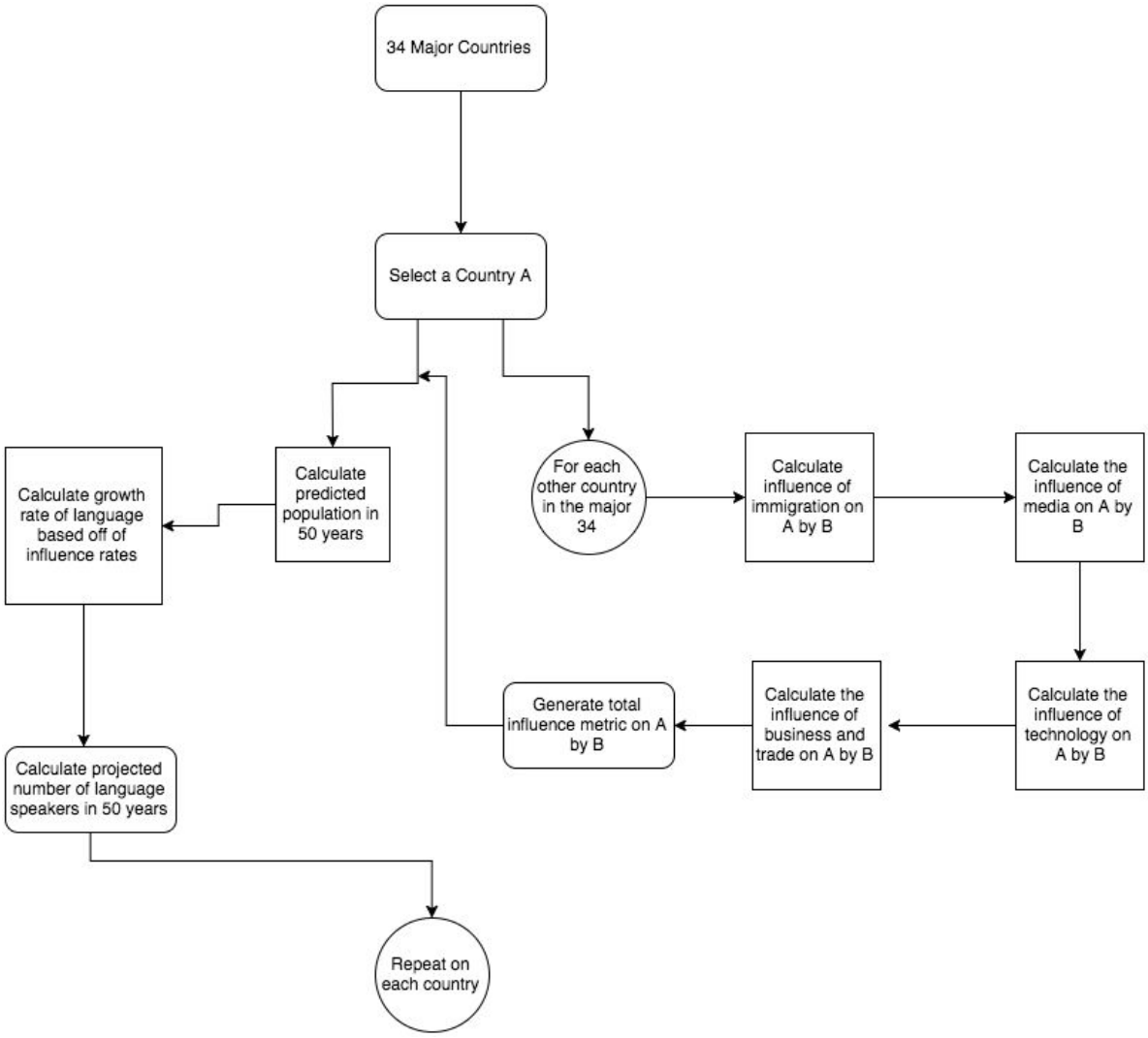


Figure 1. Model flow.

After measuring the influence rate for each country on the original, we compressed the influence rates for countries with the same dominant language(s) and determined growth rates for each language in the original country. From this, it calculated the number of language speakers in 50 years, which we then combined for each of the 34 major countries to get a distribution of the world's languages in 50 years.

3.2 Analysis of Data

Overview Our model was limited by the data that was available to us. Once we considered all the major factors that affect the growth of a language within a country, we took to the internet to figure out how to best measure each metric. In some cases, the data was thorough and readily available. In others, there were massive holes in the data that made it impossible to pursue that metric. We ended up favoring a few factors that were consistently available across all countries and gave a big picture sense of influence, and then worked within those to account for unavailable data.

3.2.1 Metrics Chosen

Population Growth Ultimately, the most important factor to consider when measuring the growth of a language is how its base population changes. Among all other factors, China's population growth has influenced Mandarin's explosion over the past century more than anything else.

Language Distribution Due to the stable nature of global language trends, the current day distribution of languages within a country was highly predictive of its future distribution.

LDI A country's Linguistic Diversity Index is a metric, compiled by SIL International and UNESCO, that measures the likelihood two people from a given population speak the same primary language. It is a good descriptor of how much crossover there is between the multiple languages that exist in a given country.

Immigration Immigration trends play a major role in how a country's population size and language distribution might change. It is important to note that, in general, when a family moves to a new country, the new country's language dominates by the third generation of that family. That meant we cared more about whether immigration was increasing or decreasing, rather than the total number of immigrants.

Trade Countries that have developed robust business relations tend to learn each other's languages. Both the exchange of goods and services demands a common language for negotiations, foreign offices, and ease of service. We wanted to give particular weight to services exchanged, but the data needed to separate trade types was not available.

Global Media Share Only a few countries are statistically significant in the global media sphere. We picked the top 10 countries that produce all kinds of media, and assumed they exerted general influence on other countries.

Technological Interaction Similarly, we measured technological influence by looking at the distribution of languages used on the internet and giving those languages general influence on each country based on how much that country uses the internet.

3.2.2 Metrics Excluded

Native vs Secondary Speakers We determined that the model would gain too much complexity if we attempted to differentiate between native languages and secondary languages. Secondary languages were included, in that the total number of language A and language B speakers is greater than the population of said country, but no special weight was given to either. Furthermore, a country's LDI will reflect its degree of language crossover.

Official Languages We did not account for internal societal factors, as doing so requires a cultural understanding of each country that is unobtainable in a weekends research. Pressures to conserve a dying language, a lingua franca that differs from the most commonly spoken language, and the languages offered in schools were not accounted for. We determined that these factors were already well reflected in the current day distribution of languages, and were generally stable.

Dialects Data on the compatibility of different dialects within the same language was largely unavailable. For some languages, the number of dialects is great and many are somewhat incompatible, such as Arabic. Unfortunately we were unable to account for this in our final model, and it is one of our largest sources of error.

Religious Influence Some religions have a high level consistency in who speaks them, like Hinduism being dominated by Hindi. Others, like Islam, have a high degree of language diversity. We determined that religion was not a major factor in learning a new language, except in minority communities, such as Jewish Americans learning Hebrew. Therefore, we did not include it in the model, as we felt its impact was dwarfed by the influence of immigration.

Tourism We initially sought to measure tourism as a reflection of country and language interaction, but determined that immigration is far more predictive of likelihood to learn a language.

Languages of Surrounding Countries In general, this metric is highly stable, and therefore is reflected in the current day distribution of languages.

Ease of Learning Language We found no metric for reliably measuring the ease of learning a language given the languages a person already knows. And because we profile by country, not individual, we determined it would be too complex to do proportional influence by every language spoken in country A to every language spoken in country B.

Economic Growth Looking at the past 50 years, economic growth was predicted closely by population growth. We decided to reduce complexity, economic growth metrics would be excluded as they are well reflected by population metrics already in the model.

3.3 Measuring Changes Over Time

Overview In order to properly measure a country's influence and its languages growth over time, we looked at changes in its historical data.

3.3.1 Metrics of Change for Influence

Delta LDI The difference in the Linguistic Diversity Index between 2017 2009 for a country. This measures the current growth of secondary languages in the country, which helps predict whether a current secondary language or new secondary language will start growing. If the Delta LDI is negative, then the dominant language(s) are becoming more dominant, which tells the model that there is a smaller chance the country becomes more multilingual. We incorporated this in the model by adding the Delta LDI of the country being influenced to the growth rate calculated for each language. This bumps up the growth rate if the country is more likely to multilingual, and brings it down otherwise.

Immigration Rate We measured the change in the rate of immigration between country A and B over a 10 year period to determine whether there is influx of country A population going to country B. The rate of change was computed relative to the total population in the country where the influx is going. This gives us the raw formula of:

$$\left(\frac{Immigration\ from\ A\ to\ B\ in\ 2015}{Population\ in\ 2015\ of\ B} - \frac{Immigration\ from\ A\ to\ B\ in\ 2005}{Population\ in\ 2005\ of\ B}\right)$$

By looking at the rate of change in immigration, we can see if going from A to B is popular for people living in A, meaning that they will be migrating with their language, which will influence the language growth in B. In addition, we calculate this relative to population in B, to see if the immigration rate actually is influencing the population, or if the population in general is just growing. Due to the generally low levels of immigration from most countries compared to the overall population, we multiplied this raw immigration rate by the constant $2 * 10^6$ as this constant gave us the most accurate results and the best spread of immigration influence among different countries. For example, before the constant was applied, all immigration rates were around < 0.01 out of 1, despite the fact that an immigration rate from Syria to Germany should be particularly high due to the influx in refugees. This was fixed with the constant.

3.3.2 Population Growth

Population growth is a large part of our model as we use it as a basis to compute the number of language speakers in a country. Using a simple logistical growth formula and the countries current rate of growth, we calculate the estimated population for the country in 50 years, and the base level of language speakers, and then we applied our growth rates for each

language to get the estimated number of language speakers. This allows for both population growth and language growth to remain logistic.

3.4 Measuring Influence

Overview The core of our model revolved around the influence of the other high-importance countries (the 34 countries we chose) on the country in question. The goal was to produce an influence metric that is between 0 and 1 inclusive, where 0 represents no influence and 1 represents very high influence. The influence metric was based off of four categories, and subsequently four other metrics.

3.4.1 Influence Metrics

Immigration Rate See Section 3.3.1

Technology The technology metric is used to measure country A's influence on country B via the internet. This is based off of two factors; country A's presence on the web, specifically their dominant language's presence on the web, and the access and use of the internet in country B. The formula used for this metric is:

$$(\%\ of\ the\ internet\ in\ language\ of\ country\ A) * (internet\ penetration\ rate\ of\ B)$$

The result of this metric were relatively low technological influences on countries. For example: The United States should have a large technological influence on most other countries, since the internet is comprised of 51% English websites. Since all of the metrics were low, we added a constant of 0.3 as an artificial boost. This makes sure the tech metric isn't ignored in the overall metric, especially since some countries with high internet penetration rates can be easily influence by a country like the United States.

Media The media metric is used to measure the influence of the exported media of country A on the media consumption of country B. Media has a large influence on people's opinions, and similarly on people's language growth, especially if they absorb media in a different language. In order to calculate this metric, we used two factors; the total share of exported media by country A, and the media consumption compared to the world average for country B. These factors allow us to generate a larger media influence metric if country A exports a lot of media and country B consumes more media than average. The formula we used for this is:

$$C * \frac{media\ exported\ by\ A\ in\ billions\ of\ dollars}{global\ media\ market\ in\ billions\ of\ dollars} * \frac{media\ consumption\ per\ person\ in\ B\ (minutes)}{average\ media\ consumption\ per\ person\ in\ the\ world\ (minutes)}$$

Once again, the result of the media metric was lower than desired, due to the fact that we are multiplying percentages against each other. Therefore,

we used a constant C, to scale the metric to an appropriate level. We decided that an influential media country such as the United States, should generally have a media metric on a random B of around or over 0.5. Using this, we set C to 2.1.

Business and Trade The last major metric we used in our influence metric was a metric to measure country A's influence on the business in country B. Business is a large driver of language growth, which is why this metric is included. In order to measure business between A and B, we looked at exports and imports between A and B. We determined that exports are more influential on the business in B, since business relationships must be formed with overseas companies to facilitate exports. Therefore, the formula is:

$$E * \%\ of\ exports\ out\ of\ A\ going\ to\ B + I * \%\ of\ imports\ going\ in\ to\ A\ from\ B$$

In order to reflect the weight of exports over imports, we created the constants E and I. We determined these constants by testing the Business and Trade metric against trade between the United States and China and between the USA and India. We set E to be 2.9 and I to be 2.3. This normalized the data and weighted the exports and imports properly.

3.4.2 Influence Metric Calculation The influence metric was put together by factoring in the four factors listed in 3.4.1. The weight of each factors, depends on the factors themselves. Specifically, when the media metric was particularly high, we determined that the influence metric should be high, even if immigration or trade was not that high. In this case, the raw influence metric looks like:

$$I = (0.1 * tech) + (0.4 * media) + (0.4 * business) + (0.1 * immigration)$$

In this scenario, immigration is weighed much lower, since it won't have a large impact on the influence due to the media overshadowing it. Business, however, still has a high impact on the influence, because there are connections between business and media, and media might affect a section of the population that might not include the business affected section. When the media metric is not high, we weigh each factor a little less:

$$I = (0.1 * tech) + (0.3 * media) + (0.3 * business) + (0.3 * immigration)$$

If the media is not overshadowing other factors in language growth, the factors should be weighed more evenly, since they each will have relatively similar effects on language growth and development. In both of these scenarios, the tech metric is weighed the lowest, since tech is correlated to media, so when tech is high, the media is often high, which

will influence the influence metric. We also weighted tech lower because websites in another language do not promote the learning of a new language, since people will just visit the website in their own language, but if a population begins to adopt a language, tech can help increase its. growth rate. After testing the influence metric, we saw that most were a little lower than expected and were generally clumped closer together than desired. For example, in the first test of the influence metric, Nigeria and Italy had a difference of .05 in their influence on the United States, despite the fact that Italy should be weighted much heavier as in influence in trade and media. After adjusting the constants for the influence metric factors, we tried different methods to widen the spread of constants, which would emphasize and assist the countries with influence, and lower the countries with basically no influence. We used a polynomial relationship to do this:

$$I = \frac{(100*I)^2}{200}$$

We found that this relationship gave us an appropriate spread for the influence metrics. However, for high influence metrics, the influence metric went above 1. So, we added a condition to bound the influence metrics to 1.

4 Model results and analysis

4.1 Places of Success

Our algorithm did well on predicting the growth of a present secondary language in a country that has a strong influence on it from another or multiple other countries. This is particularly evident in countries with English speaking populations, or where English is taught in schools, but is not the dominant language. Figure 2 shows an example of this with Russia, where currently English is spoken by 7 million people. Other examples of success in this category are China on neighboring trade partners like Japan and the United States. China has a large influence on these countries due to its influence on business and trade, as well as its presence in the international media landscape. In addition, Mandarin is expected to grow significantly in these populations, which supports this success. Our model also did well predicting the growth of a very minor language in a country with a high influence rate from other countries. This scenario is what spread English around the world, making it the Lingua Franca. This scenario is especially important, because it shows new, growing languages that will have a growing importance in 50 years, and not just languages like English and Mandarin, which will continue their growth as expected. We saw success with this for languages like Arabic, German, Spanish and Bengali. Arabic grew significantly in most countries due to the high immigration rates from Arabic speaking countries, as well as their influence in trade in the Oil and Energy industries. Arabic growth is shown in Figure 3. German grew significantly in many countries as well due to its presence as an economic powerhouse in the world. German has high influence rates over most countries it does trade with, which drove up its language use in countries like the United Kingdom and Pakistan. Spanish also had high

growth rates, which we predict is due to the media influence of Latin America as well as some trade and immigration factors. Spanish's growth was notable in Italy, as shown in Figure 5. Lastly, Bengali also saw some noticeable growth in neighboring countries due to immigration factors and developing trade. This can be seen in Figure 6.

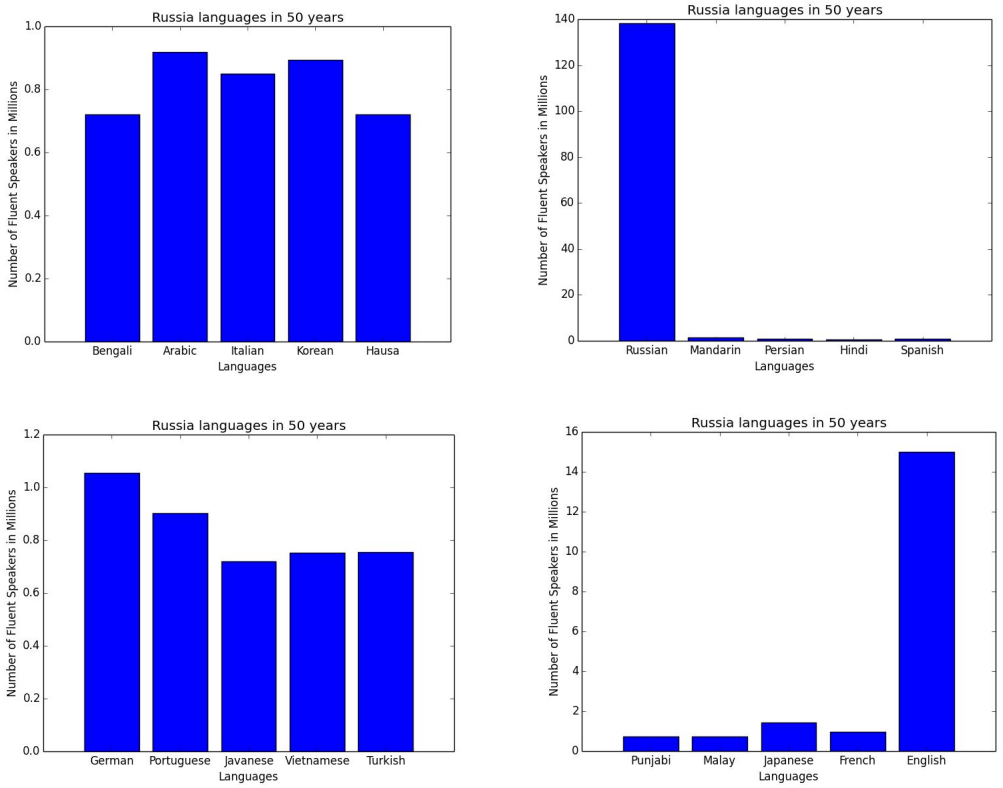


Figure 2. Growth of languages in Russia in 50 years.

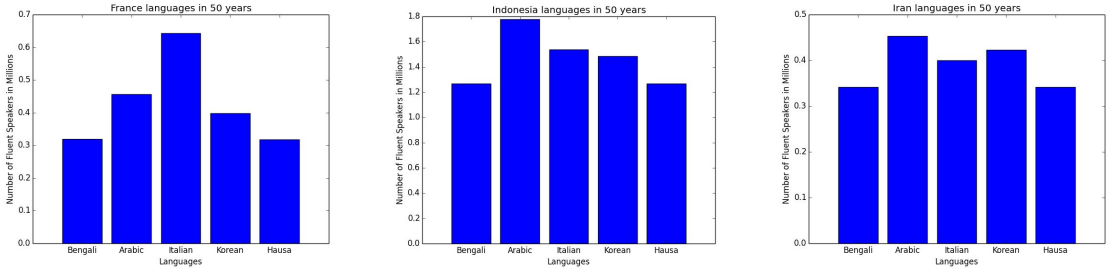


Figure 3. Growth of Arabic in various countries.

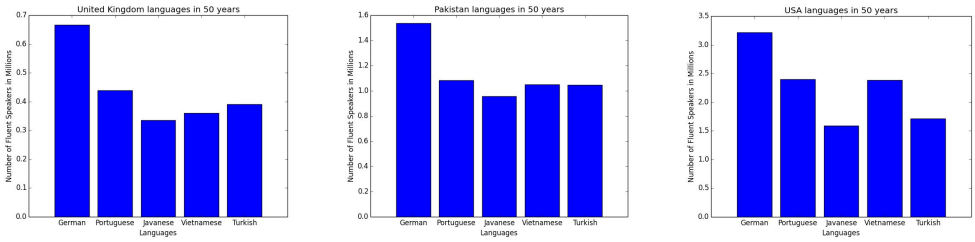


Figure 4. Growth of German in various countries.

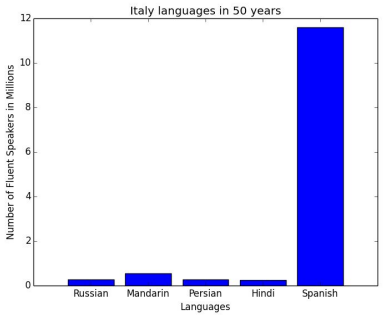


Figure 5. Growth of Spanish in Italy.

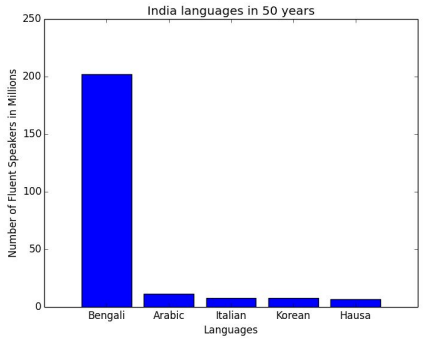


Figure 6. Growth of Bengali in India.

4.2 Places of Weakness

Our model did not do as well on less multilingual countries with most of the population speaking the same language. For example, Japan and South Korea are both dominated by their respective languages, but are surrounded by influential trade partners and parts of the global economy, so their linguistic landscape could change. However, our model kept these countries mostly one-language dominant. Similarly, our model over-grew many of the dominant languages. For example, English and Russian, in the United States and Russia, grew more than they likely will grow, due to the generated growth rate for the dominant language(s). These results are shown in Figure 7.

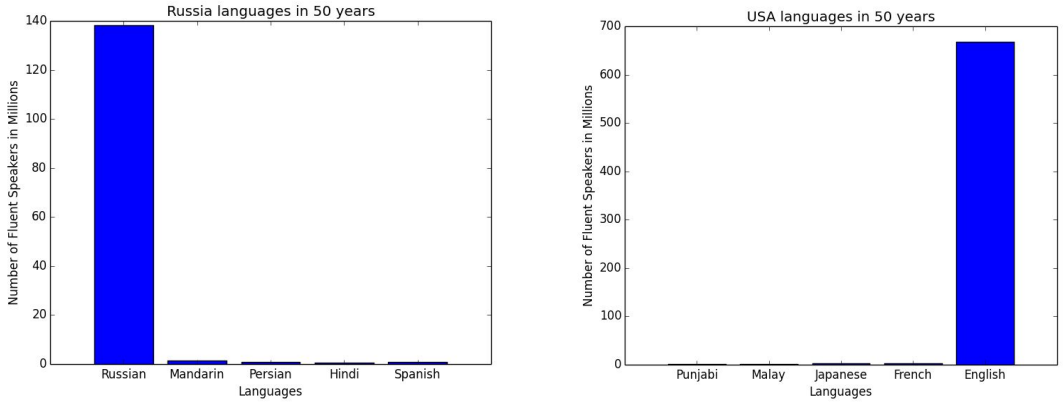


Figure 7. Growth of dominant languages.

5 Model Evaluation

5.1 Strengths

1. Our model captures every major factor that influences language growth, and applies them judiciously to develop predictions that aligns with current research and predictions. Not only do we capture large trends, but we also account for smaller trends, such as the observed increase in Arabic speakers across the globe. The main strength of our model is that these small trends are observed in many places and compiled into a larger, global prediction.

2. We also make the most of the data available by picking representative countries to describe the similar countries within their region. This gave us high coverage of the top 23 languages, and allowed us to give appropriate influence to regions with many smaller countries that would otherwise be ignored, such as Western Africa.

5.2 Sources of Error

1. Ultimately, we were limited by the data we could find. More robust information would have allowed us to make more specific evaluations of each country, which would improve the nuance of each language's influence.

2. Several recent global events skewed the data we do have. For example, when modeling immigration from Syria to Germany, with 2017 as our endpoint, we see a massive increase in the past year that makes it hard to accurately predict the future. Furthermore, refugee events tend to have permanent effects on specific regions, but are impossible to predict.

3. We concern ourselves only with fluency, as we determined that a business interested in reaching people in their own language would want to both employ fluent speakers and talk to people in their language of choice. However, despite

this decision, our model could surely be improved by weighing our metrics by primary and secondary languages.

4. Certain countries exist in specific and unusual contexts, which we did not account for. For example, slightly less than a third of Korean speakers live in North Korea, but NK behaves so differently than every other country that it does not fit in our model. Given more time and data, it would be possible to create qualifiers for every different country included in the model.

5. The direction of our research as well as our perception of what factors were most important and what results seemed reasonable was, without doubt, heavily influenced by our upbringing and respective regions of origin. We used the data and compared our results to existing research in order to reduce this bias, but ultimately everything we had access to was written by people using our preferred languages and raised in our cultures.

6 Places to Improve

1. A more specific prompt would allow us to adjust our algorithm to give greater importance to relevant factors. For example, we know the company we are delivering our results to is an international services company, but we do not know what kind of services. Luxury services have a very different target demographic than do essential services, so narrowing our model to give more weight to wealthier or less fortunate countries could have a serious impact on our recommendations.

2. We briefly discussed running the model iteratively, so instead of predicting the next 50 years at once, we would predict one year at a time. This requires modeling not only language growth but also the change in all of our other factors. However, it would give more accurate data, as well as data over time, which could be highly relevant to a company trying to open offices to hit certain trends.

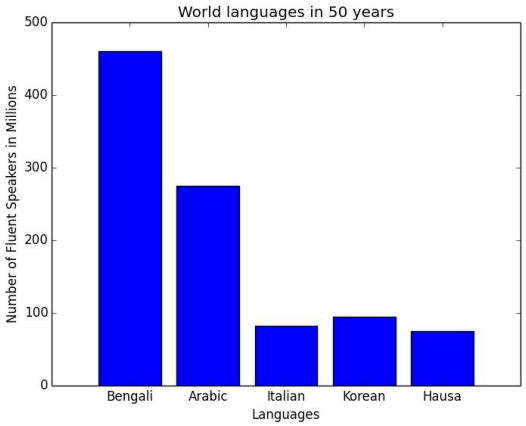
3. More data would also allow us to include every country and more languages, some of which could have a specific qualifiers applied to them. Cultural knowledge of more countries would also allow us to build a model that more accurately reflects the differences between how peoples interact with foreign cultures and languages.

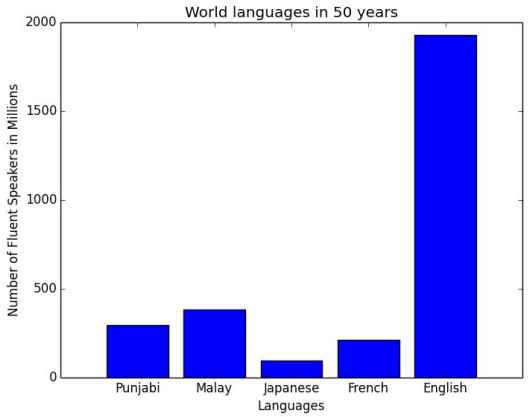
7 Conclusions

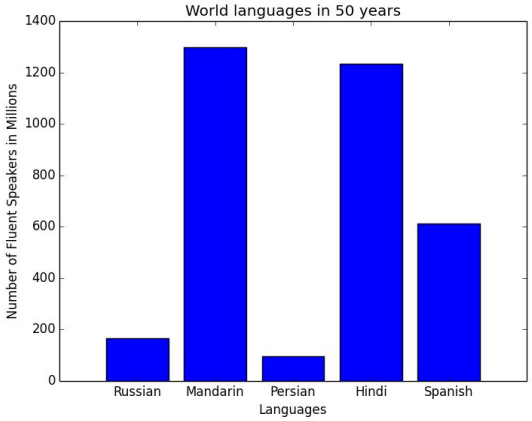
Our model for the worldwide growth of languages shows primarily an international adoption of Bengali, Arabic, English, Malay, Mandarin, Hindustani, Punjabi, French, Spanish, and Portuguese. These languages proved to dominate the more influential and populated countries and regions, continuing to spread their dominant languages to those areas that they interact with most.

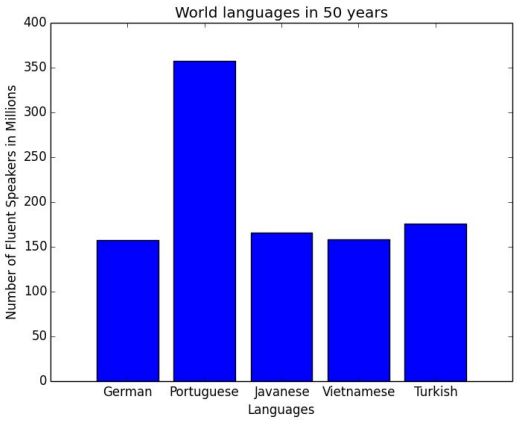
Accordingly, the main metrics that played into our model were the percentages of populations in each country that spoke a certain language fluently and that country's ability to influence the others used in our model. As previously stated, these countries were chosen since they provided a significant percentage of speakers for the world's current most popular 26 languages. Influencing other countries was determined by the formula outlined above, which primarily took into account: immigration trends, trade relationships, domestic and foreign media outlets, and the makeup of the tech industry - represented by the current breakdown of the internet based on language.

Below is a series of charts showing our predicted total speakers of the top 20 languages we identified to be dominating the world after 50 years.









Notable trends based on the above plots:

- English proceeded to overtake Mandarin Chinese as the most spoken language worldwide, most likely due to its prevalence in trade, media, and tech.
- The top three languages remained English, Mandarin Chinese and Hindustani by a significant margin, still followed by Spanish.
- Bengali leaped from the eighth most spoken language to the fifth, followed by Malay, Punjabi, Portuguese, and Arabic.
- As noted above, Punjabi rose into the seventh most popular language, while Russian fell out of the top ten.

An important demographic trend in the growth of these languages is that, although lower on the popularity of the language, Arabic and German proceeded to have positive growth almost unanimously - proving to be truly internationally adopted languages. This is an important note to make for an international company as these languages could prove useful in the targeting of worldwide markets.

Our recommendations for locations to open new offices are based primarily on the total number of speakers, density of those speakers, overlap with neighboring countries and languages, and ability to spread to other secondary markets. This is reflected in our business recommendations outlined in the memo above. Given a scenario with a stronger focus on saving company resources, we would recommend targeting the locations with more centralized dominant markets, with a smaller emphasis on international adoption of a language. Specifically, this would be implemented by targeting office locations in Indonesia, India, and Brazil, before targeting the less concentrated and defined markets in the Middle East, Germany, and Nigeria. While these later three locations were included in the memo to our COO, given an emphasis on stability and conservation, the previous three markets would probably provide safer investments.

8 Citations

Works Cited

Google. *Google*,

www.google.com/search?q=population+of+argentina+in+2010&ie=utf-8&oe=utf-8&client=firefox-b-1-ab. Accessed 10 Feb. 2018.

"Index of Linguistic Diversity." *Web Archive*,

web.archive.org/web/20120822015417/http://www.terralingua.org/linguisticdiversity/. Accessed 10 Feb. 2018.

"The Index of Linguistic Diversity: A New Quantitative Measure of Trends in the Status of the World's Languages." *Scholarspace*,

scholarspace.manoa.hawaii.edu/bitstream/10125/4474/1/harmonloh.pdf. Accessed 10 Feb. 2018.

"International Migration Database." *International Migration Database*, OECD, stats.oecd.org/Index.aspx?DataSetCode=MIG. Accessed 9 Feb. 2018.

"International Migration Statistics." *Migration Policy*, Migration Policy Institute, www.migrationpolicy.org/programs/data-hub/international-migration-statistics. Accessed 10 Feb. 2018.

"Kiswahili, a Dying Language." *The New Times*, www.newtimes.co.rw/section/read/6132. Accessed 10 Feb. 2018.

Language Projections: 2010 to 2020. *Census*, www.census.gov/hhes/socdemo/language/data/acs/Shin_Ortman_FFC2011_paper.pdf. Accessed 10 Feb. 2018.

"Languages Are Dying, but Is the Internet to Blame?" *Wired*, www.wired.co.uk/article/linguistic-diversity-online. Accessed 10 Feb. 2018.

"Languages of Bangladesh." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/List_of_languages_of_Bangladesh. Accessed 9 Feb. 2018.

"Languages of Brazil." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Brazil. Accessed 9 Feb. 2018.

"Languages of China." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_China. Accessed 9 Feb. 2018.

"Languages of Colombia." *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Colombia. Accessed 9 Feb. 2018.

"Languages of Egypt." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Egypt. Accessed 9 Feb. 2018.

"Languages of Ethiopia." *Languages of Ethiopia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Ethiopia. Accessed 9 Feb. 2018.

"Languages of France." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_France. Accessed 9 Feb. 2018.

"Languages of Germany." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Germany. Accessed 9 Feb. 2018.

"Languages of India." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_India. Accessed 9 Feb. 2018.

"Languages of Indonesia." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Indonesia. Accessed 9 Feb. 2018.

"Languages of Iran." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Iran. Accessed 9 Feb. 2018.

"Languages of Italy." *Slow Italy. Your Guide to Italy*, slowitaly.yourguidetoitaly.com/languages-of-italy/. Accessed 9 Feb. 2018.

"Languages of Nigeria." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/List_of_languages_of_Nigeria. Accessed 9 Feb. 2018.

"Languages of Pakistan." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Pakistan. Accessed 9 Feb. 2018.

"Languages of Peru." *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Peru. Accessed 9 Feb. 2018.

"Languages of Russia." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/List_of_languages_of_Russia. Accessed 9 Feb. 2018.

"Languages of the Philippines." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_the_Philippines. Accessed 9 Feb. 2018.

"Languages of the United Kingdom." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_the_United_Kingdom. Accessed 9 Feb. 2018.

"Languages of the United States." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_the_United_States. Accessed 9 Feb. 2018.

"Languages of Turkey." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Turkey. Accessed 9 Feb. 2018.

"Languages of Venezuela." *Wikipedia*, en.wikipedia.org/wiki/Languages_of_Venezuela. Accessed 9 Feb. 2018.

"Largest Media Markets Worldwide." *Statista*, www.statista.com/statistics/260065/largest-media-markets-worldwide/. Accessed 10 Feb. 2018.

"Linguistic Diversity Index." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Linguistic_diversity_index. Accessed 10 Feb. 2018.

"List of Countries by Number of Internet Users." *Wikipedia*, en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users. Accessed 10 Feb. 2018.

"List of Countries Where Arabic Is an Official Language." *Wikipedia*, en.wikipedia.org/wiki/List_of_countries_where_Arabic_is_an_official_language. Accessed 10 Feb. 2018.

"The Measurement of Linguistic Diversity." *Jstor*, www.jstor.org/stable/410659?origin=crossref&seq=1#page_scan_tab_contents. Accessed 10 Feb. 2018.

"Media and Entertainment." *Trade*. *International Trade Administration*, www.trade.gov/topmarkets/pdf/Media_and_Entertainment_Top_Markets_Report.pdf. Accessed 10 Feb. 2018.

"Media Consumption Forcasts." *Communicate Online*, communicateonline.me/wp-content/uploads/2016/06/Media-Consumption-Forecasts-2016.pdf. Accessed 10 Feb. 2018.

"Mexican Culture: Customs and Traditions." *Live Science*, www.livescience.com/38647-mexican-culture.html.

"The Relationship between Language and Careers in Multinational Corporations: A Case Study of UPM - Kymmene Oyj." *Aalto University School of Economics*, epub.lib.aalto.fi/en/ethesis/pdf/12559/hse_ethesis_12559.pdf. Accessed 10 Feb. 2018.

"The Second Most Spoken Languages Around the World." *Olivet Nazarene University*, graduate.olivet.edu/news-events/news/second-most-spoken-languages-around-world. Accessed 9 Feb. 2018.

"Tourism Statistics." *World Trade Organization*, www.e-unwto.org/toc/unwtotfb/current. Accessed 10 Feb. 2018.

"Trade Profiles." *Statistics Database*, World Trade Organization, stat.wto.org/Home/WSDBHome.aspx?Language=E. Accessed 9 Feb. 2018.

"Usage of Content Languages for Websites." *W3Techs*, w3techs.com/technologies/overview/content_language/all. Accessed 10 Feb. 2018.

"US Department of Commerce." *Commerce Data Hub*, data.commerce.gov/. Accessed 10 Feb. 2018.

"Vietnamese Language." *Wikipedia*. *Wikipedia*, en.wikipedia.org/wiki/Vietnamese_language. Accessed 9 Feb. 2018.

"What Languages Are Spoken in Japan." *World Atlas*, www.worldatlas.com/articles/what-languages-are-spoken-in-japan.html. Accessed 9 Feb. 2018.

"What Languages Are Spoken in the Democratic Republic of Congo." *World Atlas*, www.worldatlas.com/articles/what-languages-are-spoken-in-the-democratic-republic-of-the-congo.html. Accessed 9 Feb. 2018.

"The World's Languages, in 7 Maps and Charts." *The Washington Post*, www.washingtonpost.com/news/worldviews/wp/2015/04/23/the-worlds-languages-in-7-maps-and-charts/?utm_term=.cf0b3cc067aa. Accessed 9 Feb. 2018.