Ilorin Journal of Science
Volume 7, Number 2, 2020, pp. 195 – 207 (Printed in Nigeria)
ISSN: 2408 – 4840 © 2020 Faculty of Physical Sciences, University of Ilorin
https://doi.org/10.54908/iljs.2020.07.02.005

ILORIN
JOURNAL OF SCEINCE

ILJS-20-019

A Markov Chain Model for Determining Language Extinction

Ikoba[1*], N. A. and Jolayemi[2], E. T.

[1]Department of Statistics, University of Ilorin, Ilorin, Nigeria.
[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria.

Abstract

An absorbing Markov chain model used to capture the dynamics of language extinction is presented in this article. The state space consists of language ability and language inability. Maximum likelihood estimates of the intergenerational transmission probabilities were obtained over three generations and the evolution of the absorbing Markov chain was studied at generational epochs. An empirical study of some of Nigeria's indigenous languages was undertaken and it was shown that some of these languages are on a steep downward keel that may result in extinction in the absence of external intervention. Most of the languages surveyed also exhibited low levels of indigenous language literacy. Analysis of the ensuing absorbing Markov chains showed declining intergenerational transmission, with possible absorption of the chain within a few generations, in some cases.

Keywords: Absorbing Markov chain; indigenous languages; intergenerational transmission probability, time to absorption.

.

1. Introduction

Languages are the vehicle of communication among humans. Indigenous languages have been undergoing various transformations over the centuries and due to the impact of ever growing globalization, many indigenous languages are declining even to the point of extinction. There has been several modelling approaches to study the extinction of languages (Abrams and Strogatz, 2003; Kirby *et al*., 2007; Vogt, 2009). However, the Markov chain categorization of language transfer has not been explored. Markov chains provide a simple but far-reaching approach to capture the dynamics of several natural processes (Grinstead and Snell, 1997; Ibe, 2005), and could be applied to determine indigenous language decline across families within a specified time interval.

The objectives of the research are to: adapt Markov chain theory to study the trajectory of indigenous language extinction; and undertake a national survey in Nigeria to determine the

*Corresponding Author: Ikoba, N. A.
Email: ikoba.na@unilorin.edu.ng

extinction profile some indigenous languages, using the adapted Markov chain model. This study is motivated by the problem of indigenous language decline and the dearth of probabilistic models to capture this decline. The use of mathematical and statistical models to capture the dynamics of language extinction is an area of growing research. Some of the relevant researches are now presented.

Language competition between two languages was captured by Abrams and Strogatz (2003), who used an ordinary differential equation (ODE) model to explore language death. Using the proportion of language speakers over time, the model was applied to capture the decline of Scottish Gaelic in Scotland (1880-2000) and Welsh in Wales (1900-2000), and projections were made up to 2020. The Abrams and Strogatz (2003) model predicted that language competition between two languages ultimately leads to the extinction of the weaker language. However, a more realistic approach must factor multilingualism as a component of most populations.

Kirby *et al*. (2007) studied the interaction of languages with several variables including biological evolution and cultural transmission. Language acquisition by the sequence of learners was analyzed as a Markov process. The transition probabilities were based mainly on the learning algorithm used for language acquisition. Kirby *et al*. (2007) concluded that the prevalent culture in a place could significantly influence the language dynamics. The basic assumption of Kirby *et al*. (2007) was that language transfer within the Markov chain is one individual per generation.

Schulze *et al*. (2008) conceptualized a Monte Carlo simulation scheme to model birth, survival and death of languages of the world. In modelling the competition between languages, each language was characterized by a finite number of independent features, each of which can take finite values within a range of possible values. With a non-trivial probability, this feature may be changed, transferred or discarded in a random or non-random manner. The simulations were based on the entirety of languages and how they are subsumed or consume other languages; hence it is silent on the population dynamics of individual languages over time.

Wyburn and Hayward (2009) applied operations research methodology to model the interaction between two populations - one unilingual and the other bilingual. The Wyburn and Hayward (2009) model was then applied to study the language vitality for instances of modern Canada and Wales. A fundamental drawback of Wyburn and Hayward (2009) model was that a language's continued viability is the vitality of a language was tied only to the size of the active language speaking population.

Kandler (2009) proposed a reaction-diffusion model to analyze the interactions within a population which had two monolingual sub-populations and a bilingual sub-population. Demographic factors were shown to have a significant impact on language competition; with the language that was demographically more endowed enjoying greater vitality in relation to the other competing language.

Vogt (2009) explored the interaction between language evolution and demography, using analytical, agent-based analytical and agent-based cognitive modelling approaches. On the basis of the analysis, Vogt (2009) concluded that the agent-based cognitive approach produced more realistic models as shown from the accompanying simulation study.

Fernando *et al*. (2010) developed a mathematical model for language extinction in the presence of competition. The model captured the interaction between the populations of two group of multilingual speakers (the first group is a high status language, while the second group is a low status language), and a bilingual group that speak both languages. The social evolution model conceptualized by Fernando *et al*. (2010) has the drawback that it cannot capture the dynamics in a multilingual population, as it is based on the assumption of bilingualism in the population. Moreover, there seemed to be no sufficient justification for the choice of values of some parameters in the model and a way of establishing language competition.

Drawing from the assumptions of Fernando *et al*. (2010), Deka and Sinha (2016) modelled language competition and endangerment for two populations with possibility of bilingualism. Estimates of the probability of each of the possible outcomes were obtained and a Markov chain categorization of the transmission from parents to offspring was presented. The model of Deka and Sinha (2016) also possesses the drawback of Fernando *et al*. (2010) in its inability to capture multilingualism.

Ikoba and Jolayemi (2016) conceptualized indigenous language transmission as a Poisson process whose extinction profile could be studied. The model provided estimates of the intergenerational transmission probability, which reflects the proportion of children imbibing their heritage language from their parents. Declines, sometimes sharp, were observed in the intergenerational transmission probabilities for a number of languages on the basis of a survey carried out in a community in Warri, Nigeria.

Ikoba and Jolayemi (2019) viewed the decline of language ability as a Susceptible-Infectious-Removed (SIR) epidemic model using some metrics of gauging language decline. The epidemiology-oriented metrics were given a language-theoretic interpretation, and these included the basic reproduction number, time to extinction, as well as the threshold of endemicity. On the basis of some surveyed Nigerian languages and secondary input data from the 2013 Nigerian Demographic and Health Survey (NDHS), some of the surveyed indigenous languages reflected declining fortunes, with low transmission quotient from parent to children. Vogl (2019) used a random walk categorization to describe the dynamics of language invasion and language death, with various metrics of language endangerment and extinction provided.

Section 2 captures the materials and methods deployed in the research. A description of the adapted Markov chain model, some relevant assumptions and theorems are presented in the section. The results from an empirical study of some of Nigeria's indigenous languages are presented and discussed in section 3. Finally, section 4 contains the conclusion and recommendation.

2. Materials and Methods

Markov chains are discrete-time, discrete-space stochastic processes (Grinstead and Snell, 1997) that provide useful information on the dynamics of the stochastic process. A collection of real-valued, time-indexed functions called random variables ($X_n,\ n = 0, 1, ...$) forms Markov chain if the conditional probability of X_{n+1} depends only on X_n, and not on previous values (Grinstead and Snell, 1997). The specified probabilistic relation is given as

$$\Pr(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \cdots, X_1 = i_1) = \Pr(X_n = i_n | X_{n-1} = i_{n-1}) = p_{ij},$$

where p_{ij} is the transition probability from state i to state j.

The transition probabilities of the Markov chain can be represented by a square matrix, called the Transition Probability Matrix (TPM), $\mathrm{P_n}$ or P. Markov chains may be homogeneous or non-homogeneous. A homogeneous Markov chain has its TPM to be independent of the time parameter, n. This is also called a stationary Markov chain and the TPM is denoted by P. On the other hand, a non-homogeneous Markov chain has time-dependent or varying TPM, denoted by$\mathrm{P_n}$, , that is, for each of the n time-epochs observed, there is a distinct TPM, $\mathrm{P_n}$. The TPM of a homogeneous Markov chain, $P = (p_{ij})$ is a stochastic matrix (that is, each row sums to 1), and in some applications, may be doubly stochastic (each row and each column sums to 1) (Ibe, 2005).

Homogeneous Markov chain theory can be conceptualized to study the transmission dynamics of an indigenous language from parent to children. As shown by Sampson (1990) and Ibe (2005), the Markov chain categorization has several applications including the movement of unskilled labourers across several job categories, consumer choices for various goods, the job/ career choice of a son in comparison to the father's profession, the movement of individuals across regions of a country, etc. Others include testing whether the child will exhibit some of the habits of the parent, for example use of hard drugs, cigarette smoking, amorous behaviour, etc. While this language-theoretic extension of the Markov chain is seemingly different from the conventional description in that, the individual's state is conditioned on the state of his/her parent, it can still be viewed as a Markov chain with the knowledge that a parent reproduces him/herself through the offspring.

An alternative way of studying the system could be through a fractional family-based transmission model utilizing fractional flows (Leeson, 1980). For the conceptualized Markov chain model, once the decision of moving into a state has been accomplished, the previous state would play no role in determining where the process might be in future. In essence, the underlying assumption governing the dynamics of the process is that the past history is independent of the future, thus justifying the Markov property.

There are four levels of indigenous language ability, which is the capacity of an individual to communicate in his indigenous language. The four levels of indigenous language ability are: understand, speak, read, and write. The lowest level of indigenous language ability is being able to understand the language, while the highest level is the ability to write in the indigenous language.

There are two possible states in which a randomly chosen individual in the population can fall into in relation to any of the levels of language ability. He either has the indigenous language ability or he does not. Let the two possible states be 0 (cannot speak) and 1 (can speak). Define p_{ij} as the probability that a randomly chosen individual is in state j given that his father was in state i $(i, j = 0,1)$. It is the conditional probability that an individual randomly chosen in the population will have the indigenous language ability, given the status of the parent. The TPM for this homogeneous Markov chain is given as

$$P = \begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \end{array}$$

It is assumed that there was total ability of the indigenous language in the immediate past generation. This assumption is borne out of the knowledge that, prior to the advent of English language and other languages of wider communication, the indigenous language is used fully as the language of communication in the community. Therefore, for that generation and previous ones, the transmission probability is 1. The transition probabilities p_{ij} are estimated from data using the method of maximum likelihood. This entails maximizing the likelihood function (or joint probability density) subject to the constraint that $\sum_j p_{ij} = 1$.

The likelihood function for a k-state Markov chain is given as

$$L = \Pr(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i,j}^{k} p_{ij}^{n_{ij}}$$

The natural logarithm of the likelihood function (called the log-likelihood) is given as

$$\log_e L = \sum_{ij} n_{ij} \log_e(p_{ij}) \quad (1)$$

The maximum likelihood estimate of the transition probabilities are derived by maximizing the log-likelihood function (equation 1) by obtaining the first derivative of $\log_e L$ with respect to p_{ij} and equating the function to 0 (Tan and Yilmaz, 2002). Thus

$$\frac{\partial \log_e L}{\partial p_{ij}} = \frac{\partial}{\partial p_{ij}} \sum_{ij} n_{ij} \log_e(p_{ij}) = 0 \quad (2)$$

It is noted that the constraint $\sum_j p_{ij} = 1$ implies that the transition probabilities for state i of the Markov chain are not wholly independent. Thus, the partial derivative of the log-likelihood with respect to each p_{ij} will be a combination of terms with p_{ij} and $1 - \sum_{j=1}^{k-1} p_{ij}$ for the last k^{th} transition probability.

For a two-state Markov chain, there are four transition probabilities to be estimated, but with the constraint that $\sum_j p_{ij} = 1$, this reduces to two unknown probabilities. The corresponding log-likelihood function is given as

$$\log_e L = n_{00} \log_e(p_{00}) + n_{01} \log_e(p_{01}) + n_{10} \log_e(p_{10}) + n_{11} \log_e(p_{11})$$

But $p_{01} = 1 - p_{00}$, and $p_{11} = 1 - p_{10}$, and these imply that the log-likelihood can be rearranged as

$$\log_e L = n_{00} \log_e(p_{00}) + n_{01} \log_e(1 - p_{00}) + n_{10} \log_e(p_{10}) + n_{11} \log_e(1 - p_{01})$$

It is seen above that only p_{00} and p_{01} need to be estimated in order to fully specify the TPM. Hence the derivatives with respect to the specified probabilities will be derived from the log-likelihood.

$$\frac{\partial \log_e L}{\partial p_{ij}} = \frac{n_{00}}{p_{00}} + \frac{n_{01}(-1)}{(1 - p_{00})} = \frac{n_{00}(1 - p_{00}) - n_{01}p_{00}}{p_{00}(1 - p_{00})} = \frac{-(n_{00} + n_{01})p_{00} + n_{00}}{p_{00}(1 - p_{00})} = 0$$

Upon rearrangement of the derivative above, it is seen that $\hat{p}_{00} = \frac{n_{00}}{n_{0.}}$ and $\hat{p}_{01} = \frac{n_{01}}{n_{0.}}$.

The estimates $\hat{p}_{10}$ and $\hat{p}_{11}$ are also obtained in a similar manner. Hence, the maximum likelihood estimates of the transition probabilities of the Markov chain can be generalized, as presented in equation (3) below (Tan and Yilmaz, 2002).

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i.}} , \tag{3}$$

where n_{ij} represents the number of transitions from state i to state j, and $n_{i.}$ is the total number of transitions out of state i, i=0, 1.

The matrix $N = (n_{ij})$ contains the observed number of transitions between the states. Thus n_{00} is the number of persons whose parents could not speak the indigenous language and they also cannot speak, while n_{01} is the number of persons whose parents cannot speak but they (through learning in formal and informal settings) could speak the indigenous language. Similarly, n_{10} is the number of persons whose parents could speak the language but they cannot speak, and n_{11} is the number of persons who can speak the language, given that their parent can speak the language.

The chain so described by the transition probabilities (p_{ij}) can be studied using established Markov chain theory. The transition probability matrices over two generations (P and P*) can also be estimated, and trends in the dynamics of indigenous language decline over the two generations will become visible upon closer scrutiny of the two transition probability matrices.

Sufficient data on indigenous language use could be elicited for various indigenous languages via a survey using an appropriately constructed questionnaire. The data will be used to estimate the transition probabilities of language transfer over two generations for various indigenous languages. It is assumed that previous generations had total language transfer from parents to children, while for subsequent generations, the level of indigenous language ability of the respondents (understand, speak, read, write) could be used as the compartmentalization or breakdown variable to segment the population of children of such respondents based on their ability to speak the indigenous language.

The assumption was made that a parent, who does not have indigenous language ability, cannot transmit such ability to the offspring. This assumption provides the basis for studying the language situation using an absorbing Markov chain. Any absorbing Markov chain ultimately gets absorbed into any of its absorbing states (Grinstead and Snell, 1997; Ibe, 2005). On the basis of the presence of absorbing states, absorbing Markov chains have no stationary distributions.

For absorbing Markov chains, things of interest are the absorption probability (the probability of extinction of the language), time of absorption (the estimated amount of time before the language becomes extinct), and the number of steps before absorption. Under the above conceptualization, in the absence of external revitalization efforts, the indigenous language ultimately becomes extinct.

3. Result and Discussion

The data for the study are primary data obtained from an indigenous language survey conducted in some Nigerian cities in 2016. The target population was urban and semi-urban residents, as indigenous languages are fairly stable in the rural areas. The research instrument deployed was an indigenous language use questionnaire conceptualized to elicit language-use data from the respondents for up to three generations.

There were 19 questions in the questionnaire that captured various aspects of the respondent's characteristics and indigenous language abilities. A total of 606 respondents participated in the survey and these respondents in turn, provided language use data on themselves, their parents, siblings and their children. The data therefore contained aggregate information on 5,659 persons across three generations. This consisted of 1,212 parents, 3,661 children, and 786 grandchildren. The languages covered in the survey were Igbo, Yoruba, Urhobo, Edo, Esan and Isoko. Of the 606 respondents, 20 (3.3%) failed to specify their tribe, showing that there was a 3.3% non-response rate for the question on the tribe of respondents. Data processing of the questionnaire was accomplished using the IBM SPSS Statistics package, version 15.

The breakdown of the respondents, according to their language is presented in Table 1. Estimates of the transition probabilities of the respondent's ability to understand, speak, read and write in his indigenous language (P_U, P_S, P_R and P_W, respectively) are presented in Table 2. The estimated intergenerational transmission probabilities over the first generation (P) and second generation (P*) for each of the surveyed languages are presented in Table 3.

Table 1: Breakdown of respondents in the language-use survey according to their language

Language	**Frequency**	**Percentage (%)**
Igbo	144	23.8
Yoruba	56	9.2
Urhobo	121	20.0
Esan	85	14.0
Edo	48	7.9
Isoko	30	5.0
Others	102	16.8
Total respondents	586	96.7
Non-respondents	20	3.3
Overall total	606	100.0

Table 2: Transition probability matrices of language ability (Understand, Speak, Read, Write) estimated for some Nigerian languages and the overall transition probabilities.

Language	$\boldsymbol{P_U}$	$\boldsymbol{P_S}$	$\boldsymbol{P_R}$	$\boldsymbol{P_W}$
Igbo	$\begin{pmatrix} 1 & 0 \\ 0.36 & 0.64 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.35 & 0.65 \end{pmatrix}$	$\begin{pmatrix} 0.59 & 0.41 \\ 0.28 & 0.72 \end{pmatrix}$	$\begin{pmatrix} 0.60 & 0.40 \\ 0.24 & 0.76 \end{pmatrix}$
Yoruba	$\begin{pmatrix} 1 & 0 \\ 0.19 & 0.81 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.19 & 0.81 \end{pmatrix}$	$\begin{pmatrix} 0.25 & 0.75 \\ 0.18 & 0.82 \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.80 \\ 0.19 & 0.81 \end{pmatrix}$
Urhobo	$\begin{pmatrix} 1 & 0 \\ 0.50 & 0.50 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.49 & 0.51 \end{pmatrix}$	$\begin{pmatrix} 0.64 & 0.36 \\ 0.45 & 0.55 \end{pmatrix}$	$\begin{pmatrix} 0.70 & 0.30 \\ 0.36 & 0.64 \end{pmatrix}$
Esan	$\begin{pmatrix} 1 & 0 \\ 0.31 & 0.69 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.31 & 0.69 \end{pmatrix}$	$\begin{pmatrix} 0.27 & 0.73 \\ 0.31 & 0.69 \end{pmatrix}$	$\begin{pmatrix} 0.30 & 0.70 \\ 0.32 & 0.68 \end{pmatrix}$
Edo	$\begin{pmatrix} 1 & 0 \\ 0.47 & 0.53 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.43 & 0.57 \end{pmatrix}$	$\begin{pmatrix} 0.86 & 0.14 \\ 0.32 & 0.68 \end{pmatrix}$	$\begin{pmatrix} 0.87 & 0.13 \\ 0 & 1 \end{pmatrix}$
Isoko	$\begin{pmatrix} 1 & 0 \\ 0.43 & 0.57 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.47 & 0.53 \end{pmatrix}$	$\begin{pmatrix} 0.67 & 0.33 \\ 0.39 & 0.61 \end{pmatrix}$	$\begin{pmatrix} 0.67 & 0.33 \\ 0.39 & 0.61 \end{pmatrix}$
Overall	$\begin{pmatrix} 1 & 0 \\ 0.39 & 061 \end{pmatrix}$	$\begin{pmatrix} 0.82 & 0.18 \\ 0.38 & 0.62 \end{pmatrix}$	$\begin{pmatrix} 0.55 & 0.45 \\ 0.33 & 0.67 \end{pmatrix}$	$\begin{pmatrix} 0.56 & 0.44 \\ 0.29 & 0.71 \end{pmatrix}$

From Table 2, detailing the transition probabilities $(P_U, P_S, P_R, and\ P_W)$ across all the languages surveyed, a clear pattern is visible. For lower levels of indigenous language ability (understanding the language, and speaking the language), it is seen that $p_{00} = 1$ for all the languages surveyed. This value validates the assumption that only parents with the basic language ability can impart such to their offspring. Thus $\boldsymbol{P_U}$ and $\boldsymbol{P_S}$ are both absorbing chains irrespective of the language. The value of p_{11} varied between 0.50 and 0.81, as shown in Table 2. The higher values reflect those languages with sufficient intergenerational transmission of the basic language ability of understanding and being able to speak one's heritage language.

It could also be seen that the transition probability matrices $\boldsymbol{P_U}$ and $\boldsymbol{P_S}$ were similar, having entries not too different. This scenario is repeated in all the languages surveyed and leads us to conclusion that, in terms of ability to understand and speak one's heritage language, there is little difference among those that can only understand or speak the language. It could thus be inferred that, basically, an ability to speak one's language implies an ability to understand the language. However, as shown from the proportions of respondents, there is a progression from understanding to speaking to reading and finally, writing in one's language. Language understanding does not necessarily imply speaking ability, but the converse is correct in that ability to speak one's language implies language understanding.

Language understanding is the lowest language ability, while ability to write in one's language implies that one is able to read, speak and understand the language. The ability to read and write in one's indigenous language is what could be referred as *indigenous language literacy*. When the entire data on language ability is aggregated irrespective of the language of the respondent, as shown in Table 2, $\boldsymbol{P_U}$ was shown to be the only absorbing chain, as the others ($\boldsymbol{P_S}$, $\boldsymbol{P_R}$, and $\boldsymbol{P_W}$) had $p_{00} \neq 1$.

Table 3: Estimated transition probability matrices for first and second generations (P and P* respectively) for some Nigerian languages.

Language	$\boldsymbol{P}$	$\boldsymbol{P}^*$
Igbo	$\begin{pmatrix} 1 & 0 \\ 0.06 & 0.94 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.36 & 0.64 \end{pmatrix}$
Yoruba	$\begin{pmatrix} 1 & 0 \\ 0.08 & 0.92 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.20 & 0.80 \end{pmatrix}$
Urhobo	$\begin{pmatrix} 1 & 0 \\ 0.29 & 0.71 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.53 & 0.47 \end{pmatrix}$
Esan	$\begin{pmatrix} 1 & 0 \\ 0.10 & 0.90 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.31 & 0.69 \end{pmatrix}$
Edo	$\begin{pmatrix} 1 & 0 \\ 0.09 & 0.91 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.50 & 0.50 \end{pmatrix}$
Isoko	$\begin{pmatrix} 1 & 0 \\ 0.43 & 0.57 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0.47 & 0.53 \end{pmatrix}$

From the transition probability matrices for the Markov chain of language speakers over the two generations, (**P** and **P***, respectively), presented in Table 3, some insights on the language dynamics could be obtained. As could be seen from Table 3, for the Igbo language, the fraction of the second generation not having the language ability is quite higher than the corresponding value for the first generation ($p_{10}^* = 0.36,\ p_{10} = 0.06$). The matrix **P** in this case does not get absorbed even up to 150 generations, while **P*** by the fifth generation, is close to absorption. The implication of this result is that more children will be born in subsequent generations who will not have Igbo language ability. The rate of change in the probabilities of language inability over the two generations exhibited a five-fold increase, which is quite massive.

It is also observed that there was a 150% increase in the proportion of the sampled population with Yoruba language inability over the two generations, as shown in Table 3. However, the rate of absorption of **P*** in this case is slower compared to the Igbo language, as the chain is not absorbed even by the tenth generation. For the Urhobo language, an 83% increase in the proportion of those with Urhobo language inability is observed, but it took only three (3) generations for p_{10} to get beyond the absorption threshold of 0.90. It thus exhibited the steepest decline among the languages surveyed, underpinning the depth of the challenge of the survival of the Urhobo language.

More than a two-fold increase in the proportion of persons with language inability was observed for the Esan language over the two generations. However, a closer look at p_{10} and p_{10}^*, show that the proportion is still quite low compared to the other languages, as only about

31% in the second generation had the language inability. There was a four-fold increase in the proportion of those with language inability over the two generations for the Edo language. While only 9% of the sample in the first generation was unable to converse in their language, a massive 50% was observed for the second generation. This is about 450% increase from the proportion in the first generation.

The proportion of the respondents with language inability for the Isoko language for the two generations showed a two-fold increase, with the second generation exhibiting a high value of 53% being unable to converse in the Isoko language even though their parents were language speakers. It would seem that this decline in indigenous language ability permeates all the tribes surveyed. However, the decline is steeper in some of the languages than the others. The Yoruba, Esan and Igbo languages showed the lowest levels of indigenous language inability, while the Urhobo, Isoko and Edo languages exhibited the highest levels of indigenous language inability. As should be expected, the time to absorption of those chains having higher values of p_{10}^{*} is far shorter than those chains with lower values. Thus, the younger generation of Yoruba, Igbo and Esan language speakers have a lower possibility of language extinction, compared to the Urhobo, Isoko and Edo language speakers.

4. Conclusion

From the study, it is quite clear that efforts should be made to ascertain the status of many indigenous languages, as the level of intergenerational transfer from parents to children should be sufficiently high for any language to continue to be virile, population-wise. While a number of researchers (Abrams and Strogatz (2003), Fernando *et al.* (2010), etc) had ranked languages based only on their population size, an additional determinant of the status of an indigenous language should be the intergenerational transmission probability. Thus, indigenous languages can be better ranked by a combination of their population size and their intergenerational transmission probabilities. Values of the intergenerational transmission probability close to 1 indicate a language that is sufficiently safe, while values close to 0 imply the language is endangered, even with a large population size. Under such circumstances, the developed models predict that extinction is possible in the absence of external intervention through language revitalization strategies.

Acknowledgement

The authors are grateful to the anonymous reviewers who made useful and invaluable comments and suggestions.

References

Abrams, D. M. and Strogatz, S. H. (2003): Modelling the dynamics of language death. *Nature*. **424**, 900.

Deka, P. and Sinha, S. K. (2016): A new probabilistic model of language competition, endangerment and extinction. *International Journal of Computer Applications*. **142**(12), 1-5.

Fernando, C., Valijarvi, R. and Goldstein, R. A. (2010): A model of the mechanisms of language extinction and revitalization strategies to save endangered languages. *Human Biology*. **82**(1), 47-75.

Grinstead, C. M. and Snell, J. L. (1997): Introduction to Probability. American Mathematical Society, Province, RI.

Ibe, O. C. (2005): *Fundamentals of Applied Probability and Random Processes*. Oxford: Elsevier Academic Press.

Ikoba, N. A. and Jolayemi, E. T. (2016): A stochastic model of indigenous language extinction in Nigeria. *Al-Hikmah Journal of Pure & Applied Sciences*. **2**(2), 143-151.

Ikoba, N. A. and Jolayemi, E. T. (2019): Modelling language extinction using susceptible-infectious-removed (SIR) model. *Statistics in Transition New Series*. **20**(4), 71-87.

Kandler, A. (2009): Demography and language competition. *Human Biology*. **81**(2), 181-210.

Kirby, S., Dowman, M., and Griffiths, T. (2007): Innateness and Culture in the Evolution of Language. *Proceedings of the National Academy of Sciences of the USA*. **104**(12), 5241-5245.

Leeson, G. W. (1980): A projection model for hierarchical manpower system. *Journal of the Operational Research Society*. **31**(3), 247-256.

Sampson, M. (1990): A Markov chain model for unskilled workers and the highly mobile. *Journal of the American Statistical Association*. **85**(409), 177-180.

Schulze, C., Stauer, D. and Wichmann, S. (2008): Birth, survival and death of languages by monte carlo simulation. *Communications in Computational Physics*. **3**(2), 271-294.

Tan, B. and Yilmaz, K. (2002): Markov chain test for time dependence and homogeneity: an analytical and empirical evaluation. *European Journal of Operations Research*. **137**, 524-543.

Vogl, G. (2019): *Language invasion and language death*. In: Adventure Diffusion. 111-131.

Vogt, P. (2009): Modeling interactions between language evolution and demography. *Human Biology*. **81**(2), 237-258.

Wyburn, J. and Hayward, J. (2009): OR and language planning: modelling the interactions between unilingual and bilingual populations. *Journal of the Operational Research Society*. **60**, 626-636.