A Stochastic Model of Indigenous Language Extinction in Nigeria

N. A. Ikoba and E. T. Jolayemi
Department of Statistics, University of Ilorin, Ilorin, Nigeria.
E-mail: ikoba.na@unilorin.edu.ng, tejujola@unilorin.edu.ng

Abstract

The need to capture the dynamics of indigenous language decline is the main motivation for this study. Drawing from the theory of Poisson and branching processes, we conceptualize language decline as a Poisson process, and derive expressions and estimates for the intergenerational transmission probability, the probability of eventual extinction and the mean number of transfers per family. For model validation, a household survey was conducted in a community in Warri, Nigeria. Demographic information about the community was elicited through a multi-dimensional questionnaire, as well as data on the actual number of children who were able to imbibe their heritage language, across two generations. A massive demographic shift was observed in the family sizes over the two generations studied, with a steep decline in language transmission from one generation to the other. Our study also showed that a great proportion of parents in the sampled community hardly impart their language to their offspring. Our findings indicate that some indigenous languages are under serious threat of extinction if conscious steps are not taken to arrest the decline. This simple, yet powerful model can be used to ascertain the status of any indigenous language.

Keywords: Language transmission; extinction probability; Poisson process.

1 Introduction

The decline and death of languages is an established fact (Nuwer, 2014) among linguists and historians. Although languages could go extinct due to random loss of effective speakers through intergenerational transmission, the time to extinction in the absence of language revitalization measures, is of interest. The prestige of English language and other ex-colonial languages in Africa has been on the increase, possibly due to their association with modernity, technological and economic advancement, information flow and their global reach (Batibo, 2005). Many people in Africa, particularly the elite, have come to consider the ex-colonial languages as central to the economic and technological development of the continent. Even at a personal level, many parents would like to see their children speak fluent English or French or Portuguese. Many of these parents would not even mind if their children had limited fluency in their own mother tongue, as these languages are not associated with social advancement, job opportunities or the wider world (Batibo, 2005).

Understanding the process underlying language extinction is an important problem in studies of language maintenance. Currently in Nigeria, many dialects of larger language groups are even almost dead, as several factors contribute to the inability of parents to transmit these dialects to their children, coupled with the unwillingness/ lack of interest of the children to imbibe those dialects. This study therefore seeks to investigate the problem of language endangerment and extinction and propose a stochastic model to effectively capture the dynamics of language change with a view to exposing the key parameters that impact the process of language extinction. We seek to investigate this problem from the viewpoint of Poisson processes with the following objectives:

(i.) develop a stochastic model of language extinction using established principles from the theory of Poisson processes; and (ii.) undertake a survey in a Nigerian community to ascertain the level of indigenous language decline, using the developed model.

This study is motivated by the need to effectively capture the scenario of indigenous language decline, which is a result of several competing factors that affect the transmission from one generation to the next. Poisson processes theory provides a simple but powerful tool for modelling, as applied to language depletion. Poisson processes constitute the simplest and most widely applied among the class of renewal processes (Gallager, 2013). In its simplicity lies the versatility of Poisson processes. A Poisson process is a renewal process in which the interarrival times are exponentially distributed random variables. We mention a few of the papers that explored modelling language extinction. Abrams and Strogatz (2003) modelled the dynamics of language death using an ordinary differential equation (ODE) model, with one of the parameters in the model as the relative status of the language. The model is appropriate for scenarios where there is language competition between two languages.

Wyburn and Hayward (2008) applied operations research methodology in modelling the interaction between unilingual and bilingual populations, and applied the model to instances of modern Canada and Wales. Kandler (2009) proposed a reaction-diffusion model to analyze the dynamics of interactions of a population with two monolingual groups and a group that is bilingual in these two languages. The results established that demographic factors, such as population growth or population dispersal, play an important role in the competition dynamics. Vogt (2009) explored the interaction between language evolution and demography, using three different approaches: analytical modelling, agent-based analytical modelling and agent-based cognitive modelling. He concluded that the agent-based cognitive models allowed for the most detailed and realistic simulations. Anyanwu (2012) examined the pattern of interaction between the main Warri languages of Ijaw, Urhobo and Itsekiri and Pidgin English, and carried out an empirical study to establish that Pidgin English had an overriding influence on the three indigenous languages of the city of Warri, and was fast eroding the continued existence of the indigenous languages. In section 2, we present a description of the system and state some relevant assumptions and theorems. Our results are presented and discussed in section 3. Finally, our conclusions and recommendations are expressed in section 4.

2 Model Description

We assume that that the counting process of the births into a community is Poisson with parameter μ. However, due to various reasons, not all children born in the community will learn to speak the indigenous language, as the language of primary communication is another language (English or Pidgin English, in the Nigerian situation). Thus the children born into the community form the input for the process counting the number of children who assimilate the indigenous language. With probability p, a child learns his mother tongue, and with probability $1-p$, the child does not learn the language. Let the counting process of births into the community be denoted by $\{N(t), t > 0\}$ with rate μ and let $\{N_1(t), t > 0\}$ denote the process of births who acquire indigenous language ability and $\{N_2(t), t > 0\}$, the process specifying births who could not acquire indigenous language ability. We assume that the community consists of members from k families. Each birth is switched independently with probability p to $\{N_1(t), t > 0\}$ and with probability $(1-p)$ to $\{N_2(t), t > 0\}$.

It may be helpful to visualize the scenario as the combination of two independent processes. The first is the Poisson process of rate μ and the second is a Bernoulli process $X_n; n \geq 1$, where $Pr(X_n = 1) = p$ and $Pr(X_n = 2) = 1-p$. The n^{th} arrival of the Poisson process is, with probability p, tagged as type

1 arrival ($X_n = 1$) and with probability $1-p$, it is tagged as type 2 ($X_n = 2$). Let $Z_i(t)$ denote the number of children born into the ith family in the community. Then if there are a total of k families in the community, at time t,

$$N(t) = Z_1(t) + Z_2(t) + \cdots + Z_k(t) \tag{1}$$

where the $Z_i(t)'s$ are independent and identically distributed random variables. It can be shown that the resulting processes $N_1(t)$ and $N_2(t)$ are each Poisson with rates $\mu_1 = p\mu$ and $\mu_2 = (1-p)\mu$, respectively.

It is of interest to evaluate the process at certain points in time t, hence the time epochs are viewed at generational points. It may be viewed as the length of one generation for the different language speaking communities. We assume that the language is learnt through interactions at home and in the informal environment, that is, indigenous language acquisition is through intergenerational transmission. We seek to investigate the extinction pattern and the time of extinction, as well as the probability of extinction of the language in the community, given a particular rate of transmission. The ability to view independent Poisson processes either independently or as splitting of a combined process is a powerful technique for finding solutions to several otherwise less tractable problems.

The Poisson distribution has a characteristic property in relation to the binomial distribution. If the random variable X is distributed as Poisson with parameter μ and if a second random variable Y has a conditional distribution, given X, of the form Binomial (X, p), then X and $X - Y$ are independent Poisson variables with respective means $p\mu$ and $(1-p)\mu$ (Kingman, 1993). The colouring theorem as expounded in Kingman (1993), which is of huge relevance to the present study is stated below.

Theorem 2.1. : *Let Pi be a Poisson process on S with mean measure μ. Let the points of Pi be coloured randomly with k colours, the probability that a point receives the ith colour being p_i and the colours of different points being independent (of one another and of the positions of the points). Let Pi_i be the set of points with the ith colour. Then the Π_i are independent Poisson processes with mean measures*

$$\mu_i = p_i\mu, \sum p_i = 1 \tag{2}$$

The Poisson process has a number of special properties which make its use and the calculation of associated probabilities often surprisingly simple. A Poisson model is usually the simplest, and in a sense, the most random way in which to describe any particular phenomenon (Kingman, 1993). The most important feature of the Poisson distribution is its additivity (Kingman, 1993).

A Poisson process is a counting process for which the interarrival times are independent and identically distributed exponential random variables. The counting process $N(t), t \geq 0$ is said to be a Poisson proces with rate $\lambda, \lambda > 0$ if

1. $N(0) = 0$
2. $\{N(t), t \geq 0\}$ has independent increments, and
3. the number of events which occur in any interval of length t is Poisson distributed with parameter λt; that is

$$Pr\{N(t+s) - N(s) = n\} = \frac{(\lambda t)^n e^{-\lambda}}{n!}, for\ all\ s \tag{3}$$

Consider a sequence of random variables $X_0, X_1, \cdots, X_n$ where X_n represents the number of children speaking the language in the nth generation. It is assumed that the population is initiated by one individual, that is, $X_0 = 1$, and when he dies he is replaced by k individuals with probability p_k, $k = 0, 1, 2, \cdots$. These individuals behave independently and identically to the parent individual, as do those in subsequent generations. The number of indigenous language speakers in the $(n+1)^{th}$ generation, X_{n+1} is given by

$$X_{n+1} = \begin{cases} (Z_1^n + Z_2^n + \cdots Z_{X_n}^n) & if\ X_n \geq 1 \\ 0 & if\ X_n = 0 \end{cases} \tag{4}$$

The sequence $\{Z_j^n : j \geq 1\}$ are independent and identically distributed random variables, independent of X_n, and Z_j^n represents the number of offspring of the j^{th} individual in the n^{th} generation. Let $G(s)$ be the probability generating function for X_n. Then

$$G(s) = \sum_{k=0}^{\infty} p_k s^k = E[s^{X_1}] = \exp^{-\mu(1-s)} \tag{5}$$

and

$$G_n(s) = E[s^{X_n}] \tag{6}$$

$G_1 = G$, and it can be shown that, for all $n \geq 1$,

$$G_{n+1}(s) = G_n(G(s)) \tag{7}$$

and

$$G_j'(1) = \mu^j \tag{8}$$

Hence at the j^{th} generation, the expected number of persons speaking the language is μ^j. If the population size $m >> \mu^j$, then the language will become extinct if and only if $\mu < 1$. Thus if m is finite, then the population of the indigenous language speakers will be exterminated in the absence of language revitalization and sustenance programmes. It can be shown that $Z_i \sim Poisson(\mu/k)$, and this may be accomplished using the moment generating function. The distribution of the number of births in the family is therefore distributed Poisson with parameter μ/k. The trend of language transfer (intergenerational transmission) over several generations can be established when such data is available. From one language to another, there is the need to estimate μ. One way of doing this is to sample from the larger population of language speakers for the specific language under study. For each sampled family, the number of children (n_i) and the number of children in the family with the basic ability to speak the language $(x_i), i = 1, 2, \cdots, g$ is obtained. We shall use the estimators

$$\hat{\mu} = \frac{1}{g}\sum_{i=1}^{g} x_i \tag{9}$$

and

$$\tilde{\mu} = \sum_{i=1}^{g} f_i x_i \tag{10}$$

where $f_i = \frac{n_i}{\sum n_i}$ It can be shown that $\hat{\mu}$ and $\tilde{\mu}$ are both unbiased estimators of μ. However in order to determine which of the estimators is efficient, the Cramer-Rao lower bound of the variance is computed and serves as a basis for comparison of the two estimators. The respective variances of the estimators are

$$Var(\hat{\mu}) = \frac{\mu}{g} \tag{11}$$

and

$$Var(\tilde{\mu}) = \mu \sum f_i^2 \tag{12}$$

The Cramer-Rao lower bound for the variance of an unbiased estimator $\delta(x_1, x_2, \cdots, x_g)$ of the Poisson parameter μ is given by

$$Var(\delta(x_1, x_2, \cdots, x_g)) \geq \frac{1}{-gE(\frac{\partial}{\partial\mu^2} \ln p(x))} = \frac{\mu}{g} \tag{13}$$

where $p(x)$ is the probability density function of the Poisson distribution. Thus, comparing the variances of the two estimators with the Cramer-Rao lower bound, we can see that $\hat{\mu}$, which is also the maximum likelihood estimator of μ, has the least variance. It is also of interest to estimate the time to extinction of the language. If we adapt the principles of stochastic epidemic models, as shown by Andersson and Britton (2000), which is modelled as a branching process, we have the following argument: Let $\{X(t), t \geq 0\}$, be the number of individuals speaking the language at time t, $t \geq 0$, and denote by D the number of offspring of a given parent. We wish to investigate the possible extinction of $X(t)$ as t grows. First, if there are m initial individuals, then there are on the average $mE(D^j)$ individuals in the jth generation, and it is intuitively clear that the process will become extinct if and only if $E(D) \leq 1$.
For the case where $E(D) > 1$, let q be the extinction probability of the branching process and assume first that $m = 1$. Then by letting D_0 be the number of children of the ancestor, we have

$$q = \sum_{k=0}^{\infty} Pr(extinction|D_0 = k)Pr(D_0 = k) \tag{14}$$

But ultimate extinction will occur if and only if all of the (independent) branches generated by these children become extinct, hence

$$q = \sum_{k=0}^{\infty} q^k Pr(D_0 = k) \tag{15}$$

Finally, when there are m ancestors the extinction probability is given by q^m.

Next, the probability that extinction finally occurs is considered. Notice that $X_n = 0$ implies $X_{n+1} = 0$.

Define the event $A_n = \{X_n = 0\}$.$A = \cup_{n=1}^{\infty} A_n$ is the event that extinction eventually occurs. $q = Pr(A)$ is the eventual extinction probability and $Pr(A_n)$ is the probability that the extinction occurs in the nth generation. Since $Pr(A_n) = G_n(0)$, it follows that

$$q = Pr(A) = \lim_{n\to\infty} G_n(0) \tag{16}$$

Ling (2011) proved a theorem that relates to the extinction probability and is relevant.

Theorem 2.2. *The eventual extinction probability q is the smallest positive root of the equation G(s)=s.*

We can thus solve the equation

$$s - G(s) = s - e^{(-\mu(1-s))} = 0 \tag{17}$$

and obtain the smallest positive root, which corresponds to the eventual extinction probability, q. This may be computed for various languages, given estimates of μ. It may also be of interest to estimate the probable generation of extinction. Let T be the exact extinction time. Then

$$\{T = n\} = \{X_n = 0, X_{n-1} > 0\} \tag{18}$$

$Pr(X_n = 0)$ is the probability of extinction by time n and $Pr(X_n = 0) - Pr(X_{n-1} = 0)$ is the probability of extinction at exact time epoch n. Since $Pr(X_n = 0) = G_n(0)$ for all n, we have

$$Pr(T = n) = G_n(0) - G_{n-1}(0) \tag{19}$$

3 Results and Discussions

A household survey was conducted in Salem City, a community in Warri, Nigeria in which a multi-dimensional questionnaire was administered on most of the households in the community. Salem City is a multilingual community having less than 150 households. The questionnaire constructed for the survey enabled us to elicit relevant data on the language use abilities of respondents for two generations of the respondent's family, as each respondent was asked about the language ability of his siblings, as well as his children, if any. From the data elicited from the respondents, estimates of the intergenerational transmission probability $(\hat{p})$, the mean number of speakers per family $(\hat{\mu})$, and the eventual extinction probability (q) were computed. Data processing of the questionnaire was accomplished by the US Census Bureau's CSPro software program. Table 1 reflects the mean age at marriage of married respondents, according to their ethnic group. Table 2 gives the frequency distribution of the number of children of respondents' parents (f_1), number of children of respondent's parents who could speak their mother tongue (f_2), number of offspring of respondents (f_3) and number of offspring of respondents who could speak respondents' language (f_4). Table 3 contain the estimates of intergenerational transmission probability for the previous generation $(\hat{p_1})$ and present generation $(\hat{p_2})$, as well as estimates of the mean number of generational transfer of speakers of the various languages $(\hat{\mu_1}, \hat{\mu_2}, \tilde{\mu_1}, \tilde{\mu_2})$. Estimates are also provided for sub-populations of the ethnic groups (e.g. married with children; and single) and the accompanying extinction probabilities q_1 and q_2.