Original Research Article

Selection of Factors Responsible for Yoruba Language Extinction Using Feature Extraction Techniques

OI Aladesote[1*], OV Johnson[1], O Agbelusi[2] and C James[2]

[1]Federal Polytechnic, Ile-Oluji, Ondo State, Nigeria.
[2]Rufus Giwa Polytechnic, Owo, Ondo State, Nigeria.

Language is a medium of communication which shows the rich cultural uniqueness of the people. The paper extracts variables or factors responsible for Yoruba language endangerment. Three hundred and sixty (360) questionnaires were distributed to Teachers of Yoruba language in secondary schools, tertiary institutions in some parts of Ondo, Ekiti, Osun, Oyo, Ogun and Lagos States and residents who are indigenes of Yoruba speaking states. Out of 360 questionnaires, three hundred and nine (309) were administered and returned. Thus, a response rate of 85.8% was achieved. A dataset was generated from the responses. Gain Ratio and ReliefF evaluation techniques using WEKA 3.7 were used to extract factors or variables after which a threshold was set and the common factor(s) of the results were selected as significant factors to actually determine variables or factors that are responsible for Yoruba language extinction. Microsoft Excel was also used to depict the data presentation of the data collected. The result shows that Lack of commitment to indigenous language, habitat displacement, colonial legacy, devastating ferocious diseases and defective language planning are factors responsible for Yoruba language extinction. It is hereby recommended that parents should be committed to the use of Yoruba as a medium of communication at home, people should erase the thought of the legacy laid by the colonial and start using Yoruba as the major means of communication and also, there should be a standard language planning by the government of Nigeria to make our indigenous languages very compulsory for all, as one of the major requirements for admission into post primary schools.

Key words: Language Endangerment, Extraction Techniques, WEKA, Gain Ratio, ReliefF, Extinction.

INTRODUCTION

Language serves as an identity for the individual. Any individual (people) that loses his/her language has definitely a loss to his/her identity (Balogun, 2013). Yoruba is one of the major recognized languages in Nigeria but the language is gradually going to extinction. The multilingual nature of Nigeria, coupled with English Language as an official language has made indigenous language like Yoruba, very difficult to grow (Balogun, 2013). The language policy in the Nigeria constitution does not give room for indigenous languages like Yoruba to develop. Olaoye, 2007 opines that when many languages came in contact as a result of education, boundary sharing and slavery, people attempt to switch languages, mix languages because of the values attached to each of the languages. Crystal (2000) states that the English language has international influence over other languages in the world. The influence is fully responsible for the death of many languages, in which Yoruba is one. The only language for communication on the internet and Information Technology (IT) is the global English.

UNESCO (2005) reveals that the endangerment on language is a very serious problem with certain consequences, this cumulated into language loss due to military, cultural, economic, political, social crises. A language is in a stage of collapse when its speakers fail to develop it and his inability to use the language for writing and communication. The majority of the Yoruba speakers declined or refused to teach their sibling their languages. This has resulted in failure to pass the language from contemporaries to another and also from one generation to another. Darhemi (2002) and Godesborg (2007) maintain that endangerment language is a language that is no longer used as a means of communication and that such language will fade away with time. Woodbury (2012) views that languages are no longer spoken to the root, due to the fact that language has been mixed with other languages.

Corresponding Author: lomaladesote@yahoo.com

Fabunmi & Salawu (2005) point out that Yoruba language is seriously endangered and that the reasons for the endangerment includes attitude of the elite community, job opportunity and economic policies, Nigerian political factors, global Information Capitalism, etc. The researchers advise that younger ones should be encouraged to learn and speak Yoruba as a way of maintaining the language. Brenzinger & Graaf (2004) make known that Intergenerational language transmission, absolute numbers of speakers, proportion of speakers within the total population, loss of existing language domains, response to new domains and media, material for language education and literacy etc., are responsible for language extinction.

2.   THE OBJECTIVES OF THE RESEARCH WORK ARE TO:

  a) Identify factors or variables that responsible for Yoruba language extinction.
  b) Extract significant factors or variables that are responsible for Yoruba language extinction.

3.   METHODOLOGY

The existing works of the authorities in the field of Language were reviewed. The first method employed was the use of a questionnaire which consists of three parts. The first part of the questionnaire gathered respondent's personal data or background, such as age group, gender, marital status, how long the respondent has been in the southwestern part of Nigeria and if the respondent can speak fluently. The second part focused on the collection of data on factors or variables that are responsible for the Yoruba language extinction, while the respondents were asked about their views on the research topic in the third part. Respondents provided information about their views through close-ended and open ended questions on this topic. Three hundred and sixty (360) questionnaires were distributed to Teachers of Yoruba language in secondary schools, tertiary institutions in some parts of Ondo, Ekiti, Osun, Oyo, Ogun and Lagos States and residents who are indigenes of Yoruba speaking states. Out of 360 questionnaires, three hundred and nine (309) were administered and returned. Thus, a response rate of 85.8% was achieved. The responses from how long the respondent has been living in the South Western part of Nigeria in the first part and the second part of the questionnaire was used to generate a dataset in which the listed factors in the second part were used as variables in the dataset and column heading. The response from each respondent formed a record in a row of the dataset while how long the respondent has been living in the South Western part of Nigeria as a class in the last column of the dataset. The year in which the respondent has been living in the South Western part of Nigeria was rated A1 for 0 – 6 years, A2 for 7-12 years, A3 for 13 – 18years, A4 for 19 – 24 years and A5 for above 25 years. The degree of agreement on the item as a factor responsible for Yoruba language was rated using Agree (A), Strongly Agree (SA), Disagree (D), Strongly Disagree (SD) and No Comment (NC), if the respondent never encountered the situation. Gain Ratio and ReliefF evaluation techniques using WEKA 3.7 were used to extract factors or variables after which a threshold was set and the common factor(s) of the results were selected as significant factors to actually determine variables or factors that are responsible for Yoruba language extinction. Microsoft Excel was also used to depict the data presentation of the data collected. The dataset has seventeen (17) attributes that are discrete in nature. Let D be set

consisting of d data samples with n distinct classes. The expected information needed to classify a given sample is given by (Karegowda, *et al.,* 2010):

$$I(D) = -\sum_{i=1}^{n} p_i \log_2 p_i \qquad \text{.........} \quad 3.1$$

Where p_i is the probability that an arbitrary sample belongs to a class C_i and is estimated by $\frac{d_i}{d}$. Let attribute A has v distinct values. Let d_{ik} be the number of samples of class C_i in a subset D_j. D_j contains those samples in D that have value a_j of A. the expected information or entropy based on the partitioning into subsets by A, is given by:

$$E(A) = -\sum_{i=1}^{n} I(D) \frac{d_{1i} + d_{2i} + \cdots d_{ni}}{d} \qquad \text{.......} \quad 3.2$$

The information gained is given by

$$\text{Gain}(A) = I(D) - E(A) \qquad \text{..........} 3.3$$

Where E(A) is the entropy of the A and I(D) is the expected information.

The
$$\text{split info}_A(D) = -\sum_{i=1}^{n} (|D_i|/|D|) \log_2 (|D_i|/|D|)$$
.......3.4

Equation 3.4 represents the information generated by splitting the training dataset D into v partitions corresponding to v outcomes of a test on the attribute A.

$$\text{Gain ratio } (A) = \frac{Gain(A)}{\text{The split info}_A(D)} \qquad \text{..........} \quad 3.5$$

$$\text{Threshold1} = (\sum_{1}^{n} \text{Gain Ratio } (A)) - \left(mi - \left(\frac{3}{2}\right)\right) \text{.....} \quad 3.6$$

Where $\sum_{1}^{n} Gain\ Ratio\ (A)$ is the sum of Gain ratio of factors or variables under consideration, n is the number of factors or variables under consideration and mi is the minimum result of Gain Ratio.

For ReliefF

$$W_i = W_i - (x_i - nearHiti)^2 + (x_{i\_} nearMiss_i)^2 \text{.....3.7}$$

$$\text{Threshold2} = \sum_{i}^{n} \text{ReliefFresult}/n - ((mi)/(\frac{3}{2})) \qquad \text{...3.8}$$

$$SF_1 = SF_{Gain\ Ratio} \cap SF_{reliefF} \text{......3.9}$$

Where $SF_{Gain\ Ratio}$ is the significant factors from Gain Ratio while $SF_{reliefF}$ is the significant factors from relief. SF_1 is the result obtained from the intersection of $SF_{Gain\ Ratio}$ and $SF_{reliefF}$

Table 1: Variables/Factors identified as factors responsible for Yoruba language extinction

Symbol used	Factor or Variable
Interm	Intermarriage
Lcil	Lack of commitment to indigenous language use
Global	Globalization
Habit	Habitat displacement
Uceg	Unfavorable climate and economic condition
Physical	Physical wellbeing of the speakers
Desert	Desertification
Unhm	Unpredictable human migration
Hpl	High percentage of literacy
Rf	Religious factors
Med	Modernization and economic development
Cl	Colonial legacy
Dcp	Defective
Nfm	Negative perception of multilingualism
Dfd	Devastating ferocious diseases
Dnsp	Development by the native speakers and suppressive language policy
Job	Job opportunity and Economic policy

Table 4.1: Gain Ratio Result of Factors under consideration

Symbol used	Gain Ratio Result	Threshold
Interm	0.064	0.063
Lcil	0.537	0.063
Global	0.045	0.063
Habit	0.112	0.063
Uceg	0.05	0.063
Physical	0.044	0.063
Desert	0.054	0.063
Unhm	0.045	0.063
Hpl	0.06	0.063
Rf	0.066	0.063
Med	0.055	0.063
Cl	0.06	0.063
Dcp	0.067	0.063
Nfm	0.093	0.063
Dfd	0.094	0.063
Dnsp	0.072	0.063
Job	0.056	0.063

4.   RESULT AND DISCUSSION

4.1   Feature Extraction

The feasibility of the research work was demonstrated using the generated dataset on two of the existing feature extraction techniques (Gain Ratio & ReliefF).

4.1.1 Gain Ratio Result

The result of the feature extraction and the threshold is shown. Table 4.1 and Figure 4.1.interm is 0.064, lcil is 0.537, global is 0.045, habit is 0.112, uceg is 0.05, physical is 0.044, desert is 0.054, unhm is 0.045, job is 0.06, hpl is 0.066, rf is 0.055, med is 0.06, cl is 0.067, dlp is 0.093, npm is 0.094, dfd is 0.072 and dnsp is 0.056. Out of the seventeen (17) factors or variables considered, nine (9) factors (global, uceg, physical, desert, unfm, job, rf, med and dnsp) are below the threshold (0.063), meaning that these nine factors are not significant; that is, they do not really cause Yoruba language extinction, while the remaining eight (8) factors (interm, icil, habit, epl, cl, dlp, nfm and dfd) that are above the threshold (0.063) are significant; that is they are highly responsible for language extinction.
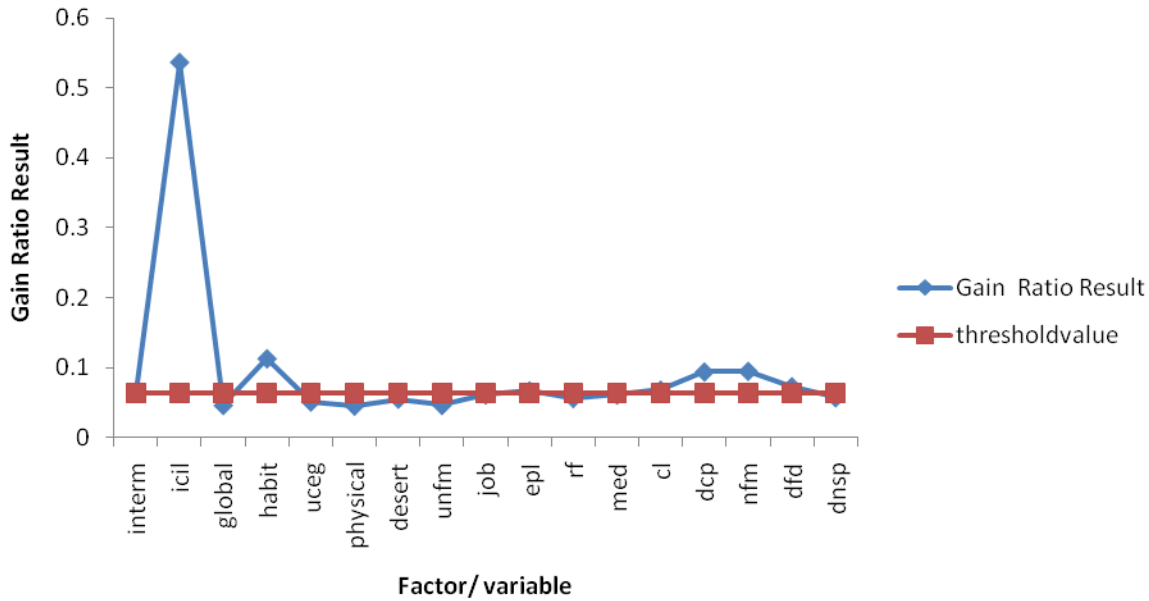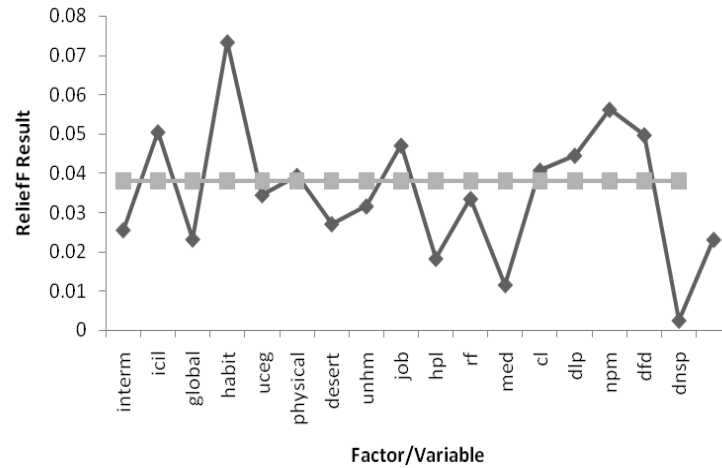
Figure 4.1: Graphical Representation of Gain Ratio Result

Table 4.2: ReliefF Result of Factors under consideration

Factor/ Variable	Gain Ratio Result	Threshold
Interm	0.0255	0.038
Lcil	0.0504	0.038
Global	0.0231	0.038
Habit	0.0734	0.038
Uceg	0.0344	0.038
Physical	0.0394	0.038
Desert	0.0270	0.038
Unhm	0.0316	0.038
Hpl	0.0471	0.038
Rf	0.0182	0.038
Med	0.0334	0.038
Cl	0.0115	0.038
Dcp	0.0401	0.038
Nfm	0.0445	0.038
Dfd	0.0562	0.038
Dnsp	0.0497	0.038
Job	0.0025	0.038

4.1.2 ReliefF Result

The result of the extraction and the threshold is shown in Table 4.2 and Figure 4.2.Interm is 0.0255, lcil is 0.0504, global is 0.0231, habit is 0.0734, uceg is 0.0344, physical is 0.0394, desert is 0.0270, unhm is 0.0316, job is 0.0471, hpl is 0.0182, rf is 0.0334, med is 0.0115, cl is 0.0407, dlp is 0.0445, npm is 0.0562, dfd is 0.0497 and dnsp is 0.0025.Out of the Seventeen (17) factors or variables considered, nine (9) factors (interm, global, uceg, desert, unhm, hpl, rf, med and dnsp) are below the threshold (0.038), meaning that these nine factors are not significant; that is they do not really cause Yoruba language extinction while the remaining eight (8) factors (lcil, habit, physical, job, cl, dlp,npm, and dfd) that are above the threshold (0.038) are significant; that is they are highly responsible for language extinction.

The common factors from the two techniques used (Gain Ratio and ReliefF) are lcil, habit, epl, cl and dlp

5. SUMMARY, CONCLUSION AND RECOMMENDATION

Language is a tribal characteristic that reveals where one comes from. It should be guided jealously to prevent it from being lost because anyone who loses his language has automatically lost his uniqueness. Language is a vital key for communication and national development. Most human activities make use of the mother tongue for communication. Man is not complete without language (Nwadike, 2008). The feasibility of the research has shown that there is little or no difference with other findings on other indigenous language extinction like Igbo. Lack of commitment to indigenous language, habitat displacement, colonial legacy, devastating ferocious diseases and defective language planning are factors responsible for Yoruba language extinction.

It is hereby recommended that parents should be committed to the use of Yoruba as a medium of communication at home, people should erase the thought of the legacy laid by the colonial and start using Yoruba as the major means of communication and also, there should be a standard language planning by the government of Nigeria to make our indigenous languages very compulsory for all, as one of the major requirements for admission into post primary schools.

REFERENCES

Balogun,T (2013). An Endangered Nigerian Indigenous Language: the case of Yoruba Language. African Nebula, Issue 6, 2013.

Brezinger, M., & Graaf, T. (2004). Documenting Endangered Languages and Language Maintenance. Available on www.mercator-research.eu/.../endangered_languages/Article%20EOLSS. Retrieved on 12 January, 2015.

Crystal, D. 2000. Language Death. Cambridge: Cambridge University Press.

Derhemi, E. (2002). Protecting Endangered Minority Languages: Sociolinguistic Perspectives Introduction". International Journal on Multicultural Societies (IJMS), Vol. 4, No. 2,150 – 161.

Godesborg, B. (2007). Language Death in West Africa. UK: Answerphone

Karegowda A., Manjunati A. & Jayaram M. (2010). Comparative Study of Attributes Selection using Gain Ratio and Correlation Based Feature Selection, International Journal of Information Technology and Knowledge Management. Volume 2, No. 2, pp. 271 – 277, July – December 2010.

Nwadike, I (2008). Igbo Langauge and Culture! Whither Bound! (Asusu na Omenala Igbo: Ije Anaa?) In R.N. Umeasiegbu (ed.) Chief (er.) F.C Ogbalu Memorial Lecturers (1 & 3). Onitsha Vasity Puiblishing Co. Ltd. In Association with Nnamdi Azikiwe University, Awka pp. 5-59.

Olaoye, A. (2013). Languages in Contact, a Blessing or a Scourge? ACase Study of Yoruba Ethnography of Greetings. Theory and Practice in Language Studies, Vol.3, No 4 pp.669 – 677, April 2013.

Woodbury, A. (2012). What is an Endangered Language? Washington: Linguistic Society of America.

UNESCO (2005). Endangered Languages. Available at http://www.mvhsmun.org/com/unesco/lang.htm. Retrieved on 21 july, 2015.