Original Research Paper Linguistic

An Understanding of Statistical Study of Language Dynamics

Rashi Agarwal	Student, SGTB Khalsa College, University of Delhi
Jasleen Kaur	Student, SGTB Khalsa College, University of Delhi
Bavneet Kaur	Student, SGTB Khalsa College, University of Delhi
Yash Mohan Sharma	Student, SGTB Khalsa College, University of Delhi
Aakash Gupta	Student, SGTB Khalsa College, University of Delhi
Tarundeep Kaur	Student, SGTB Khalsa College, University of Delhi
Sushil Kumar Singh	Assistant Professor, SGTB Khalsa College, University of Delhi
Savinder Kaur	Associate Professor, SGTB Khalsa College, University of Delhi

ABSTRACT *Language is an integral part of human thought process and has been instrumental in the development of human race. But often two or more languages have competed for speakers for various reasons. It is also true that many of the languages of the world are on the verge of extinction with no speakers. In this work we discuss the effect of language diffusion and the ensuing snowball effect initiated by economic integration just after the political independence in one of the Indian states.*

KEYWORDS : Computation Linguistics, Language Dynamics, Language Modeling.

INTRODUCTION

From the very existence of human life to present day, the development of human race has always been accompanied with the development of nature, culture and thoughts. Language forming a link with nature, human race and their culture, has always served as a basic mode of transport for thoughts' movement and evolution. Just the way every living being on the Earth, language too seems to follow the strict path of "Survival of The Fittest", as a consequence of which, the language war at present has become a major issue to ponder upon. Many languages in which histories were written are now written in history. Languages have come to inevitable collision course with one another as a consequence two or more languages have competed for speakers for various reason. It is also true that many of the languages of the world are on the verge of extinction with no speakers. Thousands of the world's languages are vanishing at an alarming rate, with 90% of them being expected to disappear with the current generation [1, 2].

Various Models have been proposed to study the language dynamics which focus on the transmission and evolution of syntax, grammar or other structural properties of a language itself [3, 4]. The flow of language is similar to that of wind, where the direction of flow depends upon the pressure difference that the words exert in the life of speaker. The gravity the words bear has always attracted the masses of different cultures. This perspective of language movement based on the internal dynamics of the language has been the focus of many studies [5].

In this paper, we bring forth some of the simple yet elegant aspects of the Diffusion Model of Language Development and apply it to the Census data of development of Hindi language in the state of Rajasthan in India. The study has much wider scope in the view that India is home to about 600 languages many of which are finding it hard to have a respectable number of speakers.

The paper is organized as follows: We introduce the Model in section 2 and in section 3 we apply it to the data obtained from the Census of India to the states of Rajasthan. In section 4, we discuss the

various issues arising out of the fitted model. Finally, we give our conclusions in last section.

THE DIFFUSION MODEL

The first model to consider would be the lexical diffusion in populations. It accounts for the spread of *words* of a given language as a consequence of learning processes [6, 7, 8, 10]. In this model it is assumed that each *word* of a given language is incorporated independently [9, 10].

Let x_i be the fraction of the population knowing the word w_i. It is evident that the rest of the population fraction $y_i = 1 - x_i$ are yet to learn the word. Now two events dominate the population dynamics of such a word. First is how fast the population fraction y_i unaware of the word learns the word

$$learning\ rate = \alpha_i y_i = \alpha_i(1 - x_i)$$

where α_i is the learning coefficient. Second is how fast the population fraction x_i unlearns the word (gives up its use).

$$extinction\ rate = \beta_i x_i$$

where β_i is the extinction coefficient. The net diffusion of the word through the population would then be

$$\frac{d}{dt} x_i = \alpha_i(1 - x_i) - \beta_i x_i$$

However, as learning is a process dependent on the interaction between individuals who already know the word with those who don't, the learning coefficient involves a linear dependence on population fraction x_i

$$\alpha_i = R_i x_i$$

Thus, it actually comes out that the population dynamics of word diffusion is non-linear and can be written as

$$\frac{d}{dt} x_i = R_i x_i(1 - x_i) - \beta_i x_i$$
$$= (R_i - \beta_i)x_i\left\{1 - \frac{R_i x_i}{(R_i - \beta_i)}\right\}$$

Let's denote $x_i^* = 1 - \frac{\beta_i}{R_i}$ then

$$\frac{d}{dt} x_i = (R_i - \beta_i)x_i\left(1 - \frac{x_i}{x_i^*}\right)$$

The equation itself gives a telltale indication that there are two equilibrium population fractions (for which the rate is null). The first is $x_i = 0$ indicative of death of the word. The other is $x_i = x_i^*$ indicative of the survival of the word [7]. Since $x_i^* = 1 - \frac{\beta_i}{R_i}$, it is also notable that the a fraction of population learn the word if the learning rate R_i is higher. It is also apparent that, for a word to be maintained in the population lexicon, the following inequality must be fulfilled $R_i > \beta_i$ [7]. The time evolution of word propagation would then be determined by the most general solution to above rate equation

$$x_i = x^* \frac{x(0)e^{(R_i - \beta_i)t}}{\{x^* + x(0)[e^{(R_i - \beta_i)t} - 1]\}}$$

THE DATA FITTING

We adopt the diffusion model as reflective of the shift of a whole collection of words and sentences. This way we apply a modified diffusion model for language shift

$$x_i = e + x^* \frac{x(0)e^{(R-1)(t-T^*)}}{\{x^* + x(0)[e^{(R-1)(t-T^*)} - 1]\}}$$

where $x_i^* = 1 - \frac{1}{R}$ with $\beta = 1$. The extra term e is to take care of the initial population fraction of the Hindi speakers. The transition time T^* signifies the time around which the snowball effect takes place.

The population fraction of Hindi speakers for the state of Rajasthan have been derived from the Census of India taking into the account the number of speakers returning Hindi as their mother tongue [11, 12, 13, 14, 15]. Table 1 lists the decadal population fraction of Hindi speakers in the state of Rajasthan for the last century. In table 2 the we have listed the best fit values of the learning rate R and the transition time T^*.

TABLE – 1

HINDI SPEAKING FRACTION FOR THE STATE OF RAJASTHAN

Census Year	Population Fraction
T	x
1911	0.1145
1921	0.1694
1931	0.1574
1941	-
1951	0.1981
1961	0.0466
1971	0.9128
1981	0.8766
1991	0.8956
2001	0.9097

TABLE – 2

THE FITTING PARAMETERS

(The adjusted-R^2 value of the fit is 0.9934.)

R	T^*
4.19	1965.75

In Figure 1 we find that modified diffusion model nicely fits the data given in table 1. The figure also displays that the fit is well within the 90% confidence level band.

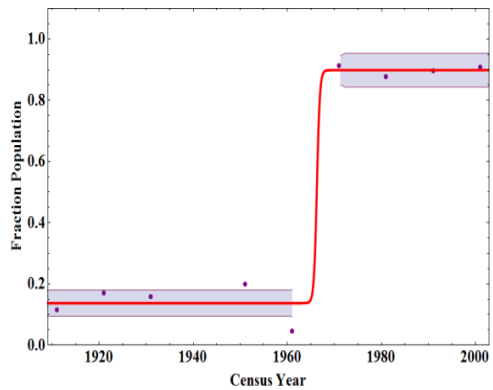

Figure 1: Fit for the Hindi speaking population over the last century with the 90% confidence level band.

DISCUSSION

The word diffusion model curve has been studied and has been shown that the word usage by fraction of population increases exponentially at low population values, describing a scenario where words rapidly propagate, followed by a slow down as the number of potential learners' decay. The accelerated, exponential growth has been dubbed as the *snowball effect* [6] and such curves have been fitted to available data [7]. In this case the economic integration just after independence of India signifies the transition time describing a scenario where language rapidly propagates in the state of Rajasthan. The pattern of Hindi language propagation is somewhat similar to the *snowball effect*. Such studies hold a lot of potential in a country as large as India having a vast treasure of languages. We strongly believe that a single model may not be enough to describe the language propagation in all states of India. Therefore, there exist a lot of scope for furthering such analysis for different languages and different parts of India.

RESULT

The modified diffusion model explains the pattern of spread of Hindi language in the state of Rajasthan during the last 100 year. Before the transition time few other major languages such as Dhundhari, Marwari and Mewari were very well competing with Hindi [11-15] but after the transition time Hindi has consumed most of the speakers of the state.

CONCLUSIONS

The study of lexicon diffusion when extended to language diffusion shows how the former drives the latter. The Figure 1 clearly indicates that apart from being gradual, the process of language propagation also involves some inevitable sudden transitions. The instant or the duration over which this transition takes place or the so called snow ball effect is observed can help us in identifying the socio-economic or natural parameters that cause the transition. The rate, at which this transition is taking place, when compared for different states, can help us tabulate these parameters in the order of their strength.

ACKNOWLEGEMENTS

The authors thank University of Delhi for financial grant under the Innovation Project SGTB-307. The authors also thank Dr Harbans Singh, SGTB Khalsa College & Dr Aranya Bhattacharjee, JNU, for their help and guidance into realms of interdisciplinary field of linguistics and physics.

REFERENCES:

[1] Krauss M. (1992), "The world's languages in crisis.", Language 68, 4–10.
[2] Nowak M.A., Komarova N.L. & Niyogi P. (2002), "Computational and evolutionary aspects of language.", Nature 417, 611–617.
[3] Hawkins J.A. & Gell-Mann M. (1992), "The evolution of human languages", Addison-Wesley, Reading, Massachusetts.
[4] Niyogi P. & Berwick R.C. (1997), "Evolutionary consequences of language learning", Ling. Phil. 20, 697–719.
[5] Wang W.S.-Y. (1969), "Competing change as a cause of residue.", Language 45, 9–25.
[6] Wang W.S.-Y. & Minett J.W. (2005), "The invasion of language: emergence, change and death.", Trends Ecol. Evol. 20, 263–269.
[7] Ricard V.S., Bernat C.M. & Jordi F. (2010), "Diversity, competition, extinction: the ecophysics of language change", J. R. Soc. Interface 7, 1647–1664.
[8] Wang, W. S.-Y., Ke, J. & Minett, J. W. (2004), "Computational studies of language evolution.", Lang. Ling. Monograph Series B 65–108.
[9] Nowak, M. A., Plotkin, J. B. & Jansen, V. (2000), "The evolution of syntactic communication.", Nature 404, 495–498.
[10] Shen, Z.-W. (1997), "Exploring the dynamic aspect of sound change.", J. Chinese Linguist. Monograph Series Number 11.
[11] Govt. of India (1961), Census of India-1961, Volume-I India, Part II C(ii), Language Table, Statement 1.
[12] Govt. of India (1961), Census of India-1961, Volume-I India, Part II C(ii), Language Table, Statement 1.
[13] Govt. of India (1981), Census of India-1981, Series-I India, Part IV B(i), Table C-7.
[14] Govt. of India (1991), Census of India-1991, Series-I India, Paper 1 of 1997, Language, India & States, Table C-7.
[15] Govt. of India (2001), Census of India-2001, Series-I India, Series-I India, Language, India, States & UT, Table C-16.