*Demography India*
*Vol. 18 (1989), pp.* 227-235

**Parameswara Krishnan**

# Linguistic Demography of India

## Introduction

INDIA is a mosaic of languages, religions, and socio-cultural groups. The Population Reference Bureau of Washington, D.C. estimates the population of India in 1989 as 835 million growing at the rate of 2.2 percent per annum. At this rate of growth, the population of India will double itself in some 32 years. The Bureau projects the population of India in the year 2000 as 1,042.5 million and in 2020 as 1,374.5 million. It may be worthwhile to state that the projected populations of China for the years 2000 and 2020 are 1291.6 and 1523.2 million. Of the projected 6.3 billion people in the world in the year 2000, India's share will be some 16% and China's 20%.

Let us look at some other socio-demographic characteristics of India. The Bureau's Population Data Sheet presents the following estimates for 1989.

Infant Mortality Rate: 96/1000 births
Crude Birth Rate: 33/1000
Total Fertility Rate: 4.3 per woman .
Crude Death Rate: 11/1000
Life Expectancy at Birth: 58 years
Degree of Urbanization: 26%
Population aged 65 years and over: 4%
Population under 15 years of age: 37%
GNP per capita: $300 (U.S.)

As stated earlier, India is a linguistic mosaic. There are inter-state differentials in the incidence of births, deaths, migration, marriages, etc. in India. The objective of this paper is to examine the linguistic, or cultural diversity factor behind these demographic phenomena.

## Literature Review

Very little is done on the linguistic sociology, or the linguistic demography, of India. Most of the studies, either with the help of census data or otherwise, have been on language

development, differences in dialects, phonology etc. Indian Linguistics talks about grammar, etimology of words and related matters. Grierson's *The Linguistic Survey of India* (11 volumes) is still the classic around which most of this type of research is done.

It was an election platform of and a commitment by the Congress Party (which has undergone change in the last two decades) to reorganize the princely states and presidencies of the Indian Union on the basis of native languages. This was done in 1956. The map of India was redrawn primarily on linguistic lines to facilitate political administration and people's participation. Though on linguistic lines, the major cities in the different States attract migrants from other States in search of job opportunities. Thus cities like Bombay, Calcutta, Delhi, Madras, Ahmedabad, Bangalore, Hyderabed all have migrants from all over India.

I.V. Sakharov (1972), in an Indo-Russian collaborative study as a part of the Census Centenary Research Program, presents the ethno-linguistic geography of the country with a commentary on some of the consequences. Lachman Khubchandani's (1972) paper on mother tongue in multilingual societies pays more attention to the Hindi-Urdu language groups and dialects. Something close to urban sociology (demography) is seen in a paper by Waheeduddin Khan and Sudhir Wanmali (1972), when the changes in the city-size distribution, as a result of linguistic reorganization, are discussed. The only demographic study that this writer has seen is a look at internal migration in India by mother tongue groupings by Ashish Bose (1973). This is included in his book on *India's Urbanization.* Since mother tongue data and other demographic characteristics are not cross-tabulated, linguistic demographic research is rendered difficult. An attempt is made here to do some preliminary work in this area.

## Population Size of Various Language Groups

The Constitution of India recognizes some 15 languages in the eighth schedule and these are all national languages. English is retained as the link language. The languages for which there are no states are Sindhi and Urdu. Hindi with its various dialects and Urdu forms the language of the largest majority (nearly 35%). In Table 1, the principal languages spoken in the various States are noted.

The estimated population of the various language (mother tongue) groups of India is shown in Table 2. Most of these groups are really large in size. Hindi and Urdu together exceed the size of the United States population. Sindhi and Kashmiri are rather small groups compared to others. The Sindhi speaking population of Pakistan is more than the one in India. The Sindhis of India are migrants from Pakistan at the time of partition of the sub-continent. Look at the size of the Bengali group which is much less in number as compared to Bangladesh (the Bureau in Washington estimates the size of Bangladesh as 114.7 million in 1989).

TABLE 1 : PRINCIPAL LANGUAGES IN VARIOUS INDIAN STATES

| *State* | *Principal Language(s)* |
|---|---|
| Andhra Pradesh | Telugu, Urdu |
| Assam | Assamese |
| Bihar | Hindi |
| Gujarat | Gujarati |
| Haryana | Hindi |
| Himachei Pradesh | Hindi, Pahari |
| Jammu and Kashmir | Kashmiri, Dadri, Dogri, Gujri, Ladakhi, Pahari, Urdu |
| Karnataka | Kannada |
| Kerala | Malayalam |
| Madhya Pradesh | Hindi |
| Maharastra | Marathi, Hindi |
| Manipur | Manipuri |
| Meghalaya | Ao, Angami, Konjak, Lotha, Sema |
| Orissa | Oriya |
| Punjab | Punjabi |
| Rajasthan | Hindi, Rajasthani |
| Sikkim | Hindi, Lepcha, Limbu, Nepali |
| Tamilnadu | Tamil |
| Tripura | Bengali, Manipuri |
| Uttar Pradesh | Hindi |
| West Bengal | Bengali |
| Andaman and Nicobar | Bangali, Hindi, Malayalam, Tamil, Telugu |
| Aranacha! Pradesh | Aka, Miji, Monpa |
| Chandigarh | Hindi, Punjabi |
| Dadra & Nagar Haveli | Bhili, Hindi, Gujarati |
| Delhi | Hindi, Punjabi, Urdu |
| Goa, Daman aand Diu | Gujarati, Konkani, Marathi |
| Mizoram | Mizo, English |
| Pondicherry | Tamil, French |

TABLE 2 : MID-YEAR ESTIMATED POPULATION OF INDIA BY MOTHER TONGUE, 1989

| *Language* | *Estimated Population (in millions)* |
|---|---|
| Assamese | 13.6 |
| Bengali | 68.2 |
| Gujarati | 39.4 |
| Hindi | 247.7 |
| Kannada | 33.1 |
| Kashmiri | 3.7 |
| Malayalam | 33.4 |
| Marathi | 64.4 |
| Oriya | 30.2 |
| Punjabi | 25.1 |
| Sindhi | 2.6 |
| Tamil | 49.1 |
| Telugu | 68.2 |
| Urdu | 43.6 |
| (Hindu + Urdu) | (291.3) |

**Note : Sanskrit was reported as mother tongue by 2000 people in 1971. English is the mother tongue of Anglo Indians**

## Comparative Linguistic Demography

Given that the states in India are on a linguistic basis, let us have an overview of the demolinguistic situation in the country. The data on crude birth rate (CBR), crude death rate (CDR) and net migration of males and females are shown in Table 3.

The differentials in the vital rates are very conspicuous. Generally, the states of Southern India have lower birth and death rates. Kerala has the lowest birth and death rate situation. States like Punjab, Haryana, Gujarat, and Maharastra have slightly higher rates of fertility and mortality in comparison to Southern India. But the States in Northern India have the highest vital rates. If we look at net migration, most of the Southern States are sending areas. The receiving States are Maharastra, West Bengal, Haryana, Punjab, Madhya Pradesh and Orissa. It may be easily seen that the net addition of people speaking other languages will result in linguistic diversity in the receiving states.

From these data, one cannot conclude that the linguistic factor operates to influence the demographic situation. The linguistic factor is closely interwoven with various socio-economic factors. We have to explore this with a proper theoretical framework.

TABLE 3 : DEMOGRAPHIC INDICATORS FOR SELECTED STATES IN INDIA

| *Slate* | *CBR 1985* | *CDR 1985* | *Net Migration ('000') 1971-81* | |
|---|---|---|---|---|
| | | | *Males* | *Females* |
| Andhra Pradesh | 29.9 | 10.3 | -237 | -215 |
| Bihar | 37.8 | 15.0 | +976 | -213 |
| Gujarat | 33.0 | 10.8 | +190 | +82 |
| Hatyana | 35.7 | 9.1 | +290 | +240 |
| Karnataka | 29.6 | 8.8 | +92 | +53 |
| KeraJa | 23.3 | 6.5 | -455 | -301 |
| Madhya Pradesh | 39.4 | 14.2 | +511 | +350 |
| Maharastra | 29.0 | 8.4 | +1,977 | +1,166 |
| Orissa | 30.7 | 14.0 | +113 | +165 |
| Punjab | 28.5 | 8.9 | +277 | +184 |
| Rajasthan | 39.7 | 13.2 | -186 | -45 |
| TamUNadu | 24.7 | 9.5 | -154 | -43 |
| Uttar Pradesh | 37.6 | 15.8 | -2,053 | -822 |
| West Bengasi | 29.4 | 9.6 | +2,821 | +1,889 |

SOURCE : Various Government of India publications.

## Theoretical Rationale

Why should one look at the language factor while doing a demographic analysis of vital events, or population processes, in a multilingual society? Language use is probably the most realistic and powerful indicator of a population's (or population group's) culture, values, and traditions. While the culture of India may be thought of as being one, there are differentials in the behavior patterns among the various language groups comprising the population, mainly arising from sub-cultural differences. Hence a 'language differential' may account for a part of the differentials in demographic behavior as well. In India's demography, the influence of Hinduism, as the dominant way of life, has been known to modify the demographic behavior of non-Hindu groups. A similar statement applies to the principal subcultures of a State as reflected in the main language of that State. Hence we can hypothesize that: the more diverse the culture of a State, the more divergent the demographic behavior in that State from the all India average level. The linguistic (and hence cultural) diversity of a State is influenced by various socioeconomic and political forces. We can see that these forces, interalia, affect the demographic situation as well. Hence to test this hypothesis, we have to resort to a multivariate statistical procedure and separate out the influence of linguistic diversity (hence cultural diversity) controlling for other factors.

Hypotheses

1. State fertility differentials are, inter alia, due to sub-cultural (language) differences. The more diverse a state in its language resources, the lower its fertility level.
2. The more diverse a state in its language resources, the lower its mortality level.
3. The more diverse a state in its language resources, the higher its net migration level.

Data and Methodology

The data employed for the analysis refer to the various states (aggregate data) for the most recent years. Some of the correlates refer to the 1981 census. Since the aim of this study is to elicit a pattern of relationship between language diversity and demographic behaviour, the different reference points in time to the various data pieces are unimportant. The unit of analysis is the state, drawn up on linguistic basis. Our dependent variables are crude birth rate *(CBR)* and crude death rate *(CDR)* indicators of fertility and mortality conditions. The social correlates utilised in the analysis are:

(a) Degree of urbanization *(URB)* measured by per cent population living in urban areas
(b) Female literacy rate *(FLR)*
(c) Female labour force participation in non-agricultural occupations *(FLFPR)*
(d) Effective contraceptive protection *(ECP)* indicated by percent women effectively protected by contraception
(e) Language non-diversity *(LAND)* portrayed by percent with the state language as mother tongue living in a State
(f) Female age at first marriage *(AGE)*

This is the singulate mean at first marriage computed from the census data on proportion single using the Hajnal procedure.

The number of data points is 29 representing the number of states and union territories in the country.

Fertility Modelling

The models for fertility are given by:

*Multiple Regression -*

$$E(CBR) = a_o + a_1^{(-)} CDR + a_2^{(-)} AGE + a_3^{(-)} URB + a_4^{(+)} LANG + a_5^{(-)} FLPR + a_6^{(-)} FLR + a_7^{(-)} ECP$$

The directions hypothesized for the regression coefficients are shown in parentheses.

***Cobb-Douglas Function***

$$E(lnCBR) = a_{01} + lnCDR + a_2^{1(-)} lnAGE + a_3^{1(-)} lnURB + a_4^{1(+)} lnLANG + a_5^{1(-)} lnFLPR + a_6^{1(-)} lnFLR + a_7^{1(-)} lnECP$$

Two types of multivariate regres, on models are used—(1) a straight multiple regression model, and (2) a Cobb-Douglas function model. Both models lead to similar results. Hence only the multiple regression model is discussed below.

The means and standard deviations of the various variables are shown in Table 4 as also the intercorrelation coefficients. The correlation coefficients between *CBR* and *CDR* and the other correlates *FLR, URB, FLPR, ECP* and *AGE* are all in the expected directions. The correlation between *CBR and LANG* is negative and negligibly small and not in the expected direction.

TABLE 4 : MEANS, S.D.'s AND INTERCORRELATION, INDIA (CIRCA 1985)

**A. Means and S.D.'s**

| | *Mean* | *SD.* |
|---|---|---|
| CBR | 31.08 | 5.51 |
| CDR | 9.84 | 3.19 |
| FLR | 30.70 | 14.44 |
| LANG | 62.20 | 28.56 |
| URB | 27.38 | 21.79 |
| FLFPR | 28.71 | 25.37 |
| ECP | 32.75 | 15.83 |
| AGE | 18.43 | 1.22 |

Number of cases - 29

B. Correlation Matrix

| | *CBR* | *CDR* | *FLR* | *LANG* | *URB* | *FLFPR* | *ECP* | *AGE* |
|---|---|---|---|---|---|---|---|---|
| CBR | 1.000 | .723 | -.708 | -.002 | -.315 | -.385 | -.275 | -.557 |
| CDR | | 1.000 | -.790 | -.017 | -.512 | -.602 | -.131 | -.418 |
| FLR | | | 1.000 | 0.092 | .656 | .758 | .253 | .551 |
| LANG | | | | 1.000 | .173 | .057 | .568 | .211 |
| URB | | | | | 1.000 | .780 | .280 | .061 |
| FLFPR | | | | | | 1.000 | .200 | .255 |
| ECP | | | | | | | 1.000 | .405 |
| AGE | | | | | | | | 1.000 |

SOURCE : Regression Analysis.

The results of a stepwise regression are shown in Table 5. This allows for the entry of all the correlates. We see that all these correlates together explain 67.7% of the variation in CBR (adjusted R2 is 58.9). But the variables AGE, URB, LANG, ECP, FLFPR, are not significant at all. Even FLR attains significance only at 10% level. When the insignificant variables are removed from the analysis, we see that the first to drop out is URB (accounting for 1% reduction), LANG (2% reduction) and ECP (about 2% reduction). Thus the parsimonious fertility model is:

$$CBR = 28.82 + .7567\,CDR + .0779\,FLFPR - .2419\,FLR$$

*accounting for 62.8 percent of the variance (adjusted* $R^2$ *being 0.5831).*

*TABLE 5 : REGRESSION OF CBR ON CORRELATES*

| | Reg. Coef. | SigT |
|---|---|---|
| *AGE* | *-.515* | *.5512* |
| *URB* | *.018* | *.7666* |
| *LANG* | *.032* | *.2814* |
| *CDR* | *.786* | *.0362* |
| *ECP* | *-.0.74* | *.2209* |
| *FLFPR* | *.060* | *.2481* |
| *CONSTANT* | *-.192* | *.1055* |
| *FLR* | *36.919* | |

$R^2 = .6787$, *Adjusted* $R^2 = .5716$

LANG, though attaining significance only at 28%, has a positive relationship as hypothesized. Note the correlation coefficient with CBR was negative. Hence the sub-cultural factor has to be recognized in studying inter-state fertility differentials in India. This is, the more 'non-diverse' a state, the higher will be its fertility level.

***Mortality Modelling***

Again two models were tried to develop an explication of inter-state mortality. The models are similar to the ones tried for fertility. The results were similar. Hence the Cobb-Douglas function model only is presented here

$$E(lnCDR) = b_0 + b_1^{(+)} lnAGE + b_2^{(-)} lnURB + b_3^{(+)} lnLANG + b_4^{(+)} lnCBR + b_5^{(+)} lnFLFPR + b_6^{(-)} lnFLR$$

The expected directions are shown in parentheses.

Most of the inter correlations are in the expected direction. When all these variables are entered, some 72.1% of the variation in *IhCDR* is explained (adjusted $R^2$ 62.8%). But only two variables turn out to be significant viz *In FLR* and *InCBR*. *InLANG* does not have any explanatory power at all. The Cobb-Douglas equation in the log scale is seen as:

$$InCDR = 2.1413 - 0.4736\ InFLR + 0.4033\ InCBR$$

The implication of this analysis is that sub-cultural factor, as reflected in language use, is not helpful to explain mortality variation at all.

It would have been worthwhile to try a simultaneous equation model. Our present data sources do not permit the use of that methodology for this paper.

**Conclusion**

The influence of culture (or subculture) on fertility and mortality has been noted in many case studies in India. In the empirical analyses, it was found that the role of the language factor in influencing either inter-state fertility or mortality differentials is rather insignificant. These are several reasons for the non-revelation of the importance of language in the data analysis. First we have relied on ecological relationship. Ecological correlation cannot be equated to the actual association that exists between an event and its correlates. Since we did not have data on the social correlates and fertility/mortality indicators at the individual level, we had resorted to the ecological analysis. Field studies have to be carried out to collect detailed micro level data to test the language influence hypotheses. The border districts between states are good areas to do some exploratory research. The various migrant families living in cities also need to be studied to see how assimilation to the local state culture is proceeding.

References


Bose, Ashish, 1973, *India's Urbanization 1901-2001,* Tata-McGraw Hill, Bombay (second edition).

Khan, Waheeduddin and Wanmali, Sudhir, 1972, Impact of linguistic reorganisation of States on city-size distribution in penisularIndia, in *Economic andSocio-cultwalDimensions ofRegionalisation,* An Indo-USSR Colloborative Study, Census Centenary Monograph No. 7, Census of India 1971, Office of the Registrar General, New Delhi, 451-472.

Khubchandani, Lachman, 1972, Mother tongue in multilingual societies, in *Economic and Socio-Culiural Dimensions of Religionalisation,* Op. cit, pp. 427-450.

Sakharov, I. V., 1972, Ethno-linguistic geography of India: facts and problems, in *Economic and Socio-Cultural Dimensions ofRegionalisation,* Op. cit., pp. 387-426.