A note on understanding the relationship between linguistic diversity and biodiversity in the Indian context

Sibansu Mukherjee

Linguistic diversity (*LD*) and biodiversity (*BD*) are related to each other[1,2,3,4,5]. A language as far as its vocabulary concerns contains indigenous knowledge that includes local terms used to describe the facts on nature, livelihood and other subsistence of the linguistic community[3]. The relationship between *LD* and *BD* has been demonstrated in different studies[3,4]. The way mainly followed in these studies is to show the congruence between language and species richness. Nevertheless, this way is complex. Instead, I demonstrate the said relationship in the Indian context with a simpler method. Here, I prefer two easier parameters as two variables: (i) population of the languages apart from the state's scheduled languages out of total population of the state[6] (call this the 'population of other languages,' *PoL* in short) and (ii) state's forest area *FA* out of the state's total area[7]. Using *Pearson's r*, I find a strong positive correlation between the two consecutive variables. Since comprehensive data on species richness, BD wilderness and LD are not always available for every area on this earth, the variables mentioned above are most useful and easily available to estimate the LD-BD relationship.

Diversity among the human groups is itself a biological phenomenon. *LD* has been considered as a result of human flexibility in learning different cultures through biological adaptation. *LD* also has an evolutionary biological basis, where genetic structure and linguistic structure are considered to be co-evolved[5,8]. Researchers[3] finding a significant congruence of high BD wilderness areas and LD proclaim that this fact is true for 70% languages of the world. Furthermore, the relationship between species richness and language variations spatially varies and is complex. They examine some high BD wilderness regions that cover 6.1% of Earth's total land area. This robust study has been done on the assumption that language contains information on natural and biological worlds, i.e., environment, plant and animal lives. Survival of humans in such regions depends on the conservation of natural and cultural environment. Moreover, this study[3] claims that the results derived from the datasets developed to explore the co-occurrence of LD and BD in certain regions can be considered as an initiation to focus research on the relationship between LD and BD and to develop integrated strategies to protect species and languages exist in such regions. Link between language diversity and species richness in certain area has also been studied where a negative correlation between threatened languages and threatened mammals is found in New Guinea[4,9].

This short note also subscribes the issue of *LD-BD* relatedness but in a different way. On the contrary to the above assessments on BD wilderness and specie richness related to *LD*, I use *FA* that exhibits comparative data on forest coverage areas in the different Indian states and *PoL* as LD indicator of the state. I find a strong positive correlation score between *PoL* and *FA* and the score implies that there is a robust relationship between *LD* and *BD*.

Why did I consider *PoL* and *FA* as two variables to specify the link between LD and BD in Indian context?

Instead of collecting data on the parameters of LD and BD used in the previous studies I have depended on available data provided by the government of India. In order to calculate co-relation between two variables, a common basis was required. I did not find such type of common basis for the variables. I got the list of state-wise languages from the website of census data, but there is no such data available on state-wise BD wilderness or state-wise species richness. Therefore, I have chosen on one hand, state-wise data on *PoL* used as the LD indicator (dependent variable) and on the other hand, state-wise data on FA used as the BD indicator (independent variable) in this study. I have developed a dataset for 12 Indian states and union territories.

India is a plurilingual country. Quantification of such plurilingualism is not possible. As a result, it is also difficult to measure how an area linguistically varied. So I had to depend on census data provided by the Government of India. In 2011, according to the principle of being *faithful to the respondent*, Indian census has reported that the number of *raw returns* of mother tongues of the country is *19569*[6]. After a methodical *linguistic scrutiny*, *edit* and *rationalization* of such raw return the number of *rationalized mother tongue* is *1369*, whereas 1474 names are treated as *unclassified*. Furthermore, the number of languages with *10000 speakers* is *121*[6]. 22 among these 121 languages are *scheduled as national languages* of India and rest 99 languages are considered mainly as *PoL*[6].

Population of 22 scheduled languages has occupied 96.71% of the total population of India whereas *other classified languages* (1369 -22 = 1347) are only 3.29% of the total population. It is very easy to understand that this 3.29% linguistic population has enormous diversity. So when I considered state-wise distribution of the population of other languages in terms of the percentage of the total population, I am basically looking at the linguistic diversity of the state.

In contrast, India as a *BD* area has been divided into 10 major *biogeographic zones* including *Marine Influenced Area*[7]. These are as follows: Trans Himalaya, Himalaya, Dessert, Semi-arid, Western ghats, Deccan peninsula, Gangetic plain, coasts, north-easts and Islands[7]. Therefore, neither a *biogeographic zone* is equal to a state nor I could have been able to collect data on state-wise species richness or biological wilderness. But, since forests are always considered as having species richness and it covers only a very small portion of the total area of the state exactly as same as the *PoL* out of the state's total population; I took *FA* as the indicator of the *BD*.

Dataset developed on *PoL* and *FA* has been developed to show the possible linkage between *LD* and *BD* as in *Table 1*. North-east India eventually is considered as having a high level of BD[10]. I also have observed that the north-east Indian states, namely, Arunachal Pradesh, Nagaland, Mizoram, Meghalaya, still have a great deal of forests and consequently a huge number of linguistic varieties. On one hand, according to 2011 census, these states have 72.13%, 88.13%, 87.65% and 85.35% population of the languages other than the scheduled languages out of the state's total population respectively. On the other hand, forests cover respectively 80.30%, 78.21%, 88.93% and 76.76% of the total area of the state. Other areas like the state Sikkim and the union territories Andaman & Nicobar Islands have statistically medium level *LD* and *BD* in my sense. Rajasthan, Jharkhand and Maharashtra also have several tribal communities and forests. On the contrary, West Bengal, Kerala and Tamil Nadu are considered as the most developed states in India and thus have very poor in terms of *LD* and *BD*. A comparative view of the data has been presented at *Figure 1*.

As per the dataset displayed at *Table 1*, I find relationship between two variables, *PoL* and *FA*, significant ($p < 0.05$) with a strong positive correlation (*Pearson's r = 0.9021*) and a high coefficient of determination ($r^2 = 0.8138$).

To examine this result, I also provide the score derived by another correlation calculation, where I have taken decadal growth rate (2011) of *rural population*[11] of these 12 states and union territory as independent variable. Presumably, the scheduled languages are related to the state's primary developmental policies and are mainly urban-centric[12].

Indian LD has been observed through different socio-cultural viewpoints[12,13]. For example, India as a part of south-east Asia has been described as a habitation of plural societies, where speaker's identity as a

language user cannot be universally defined. Thus, so-called language boundaries in India remain fuzzy and fluid[13].

The consciousness about language identity usually grows among the urban population. Most of the rural societies do not concern about their language identity. So far is studied that the rural areas have high level of linguistic diversity. As there is no such model to quantify different level of LD, I have tried to use the decadal growth (2011) rate of rural population instead of FA. Table 2 has flourished dataset where Independent variable has been changed.

Here, the value of *Pearson's r* is *-0.9339*. Interestingly, relationship between rural population and the speaking population other than the scheduled languages shows a strong negative correlation but the coefficient of determination is as strong as previous (r^2 = *0.8722*). A simple assumption I have taken for this. A language with a huge number of speakers is usually considered in the 8th schedule of the Indian census as national language. A language having such status seems to have a great policies and planning that satisfy state's main agenda. As a result, non-scheduled languages remain lesser planned and in most of the cases unplanned and thus diverse. Both linguistic and biodiversity need to be protected for the sake of Earth's health.

World is facing a controversial human move. On one hand, researchers are worried on the issues of language extinction considerably linked with species extinction. On the other hand, humans irrespective of ethnicities cultures professions are allowed to participate in the global electronic venture that only provide a common hybrid platform of language. This accountability of the relationship between languages other than the standard planned languages and the land areas covered by forest provides us thinking about the necessity of conservation of forest for the sake of biodiversity as well as *LD*. the strong positive correlation between *PoL* and *FA* proves the claim of relatedness between *LD* and *BD* is justified and much intense studies referred earlier on language and species richness are path breaking. Moreover, this study also provides an alternative POV as well as method to account relationship between *LD* and *BD*, especially when statistical data on languages and species richness is unavailable.

References

1. Skutnabb-Kangas, T. (NJ & UK: LEA. 2000)
2. Upadhyaya, R. K. & Hasnain, S. I. *Lingua* 195, 110-123, https://doi.org/10.1016/j.lingua.2017.06.002 (2017).
3. Gorenflo, L. J. et al. PNAS 109 (21), 8032-8037, https://doi.org/10.1073/pnas.1117511109 (2012).
4. Turvey, S.T. & Pettorelli, N. *Proc.R.Soc. B* 281, 20141644, doi:10.1098/rspb.2014.1644 (2014).
5. Baronchelli, A. et al. *PLoS ONE* 7(10): e48029. Doi:10.1371/journal.pone.0048029 (2012).
6. Census of India 2011. *Language* (paper 1 of 2018), (ND: Office of the Reg. Gen. India. 2018)
7. GoI. *Compendium of Environment Statistics*. (ND: Central Statistics Office. 2015)
8. Hammarström, H. *Jour. of Lang. Evol.*, 1(1), 19–29, https://doi.org/10.1093/jole/lzw002 (2016).
9. Cardillo, M. et al. *Proc. R. Soc. B*. DOI: 10.1098/rspb.2014.2986 (2015).
10. Singh, E. J. et al. NeBIO 1(1), 42-47, (2009).
11. Rao, K. H. Rural development statistics. (Hyderabad: NIRD. 2012).
12. Gumperz, J. J. 1957. Language Problems in the Rural Development of North India. The Journal of Asian Studies, 16, pp 251-259 doi:10.2307/2941382
13. Khubchandani, L. M. (ND: Sage. 1997)

Tables

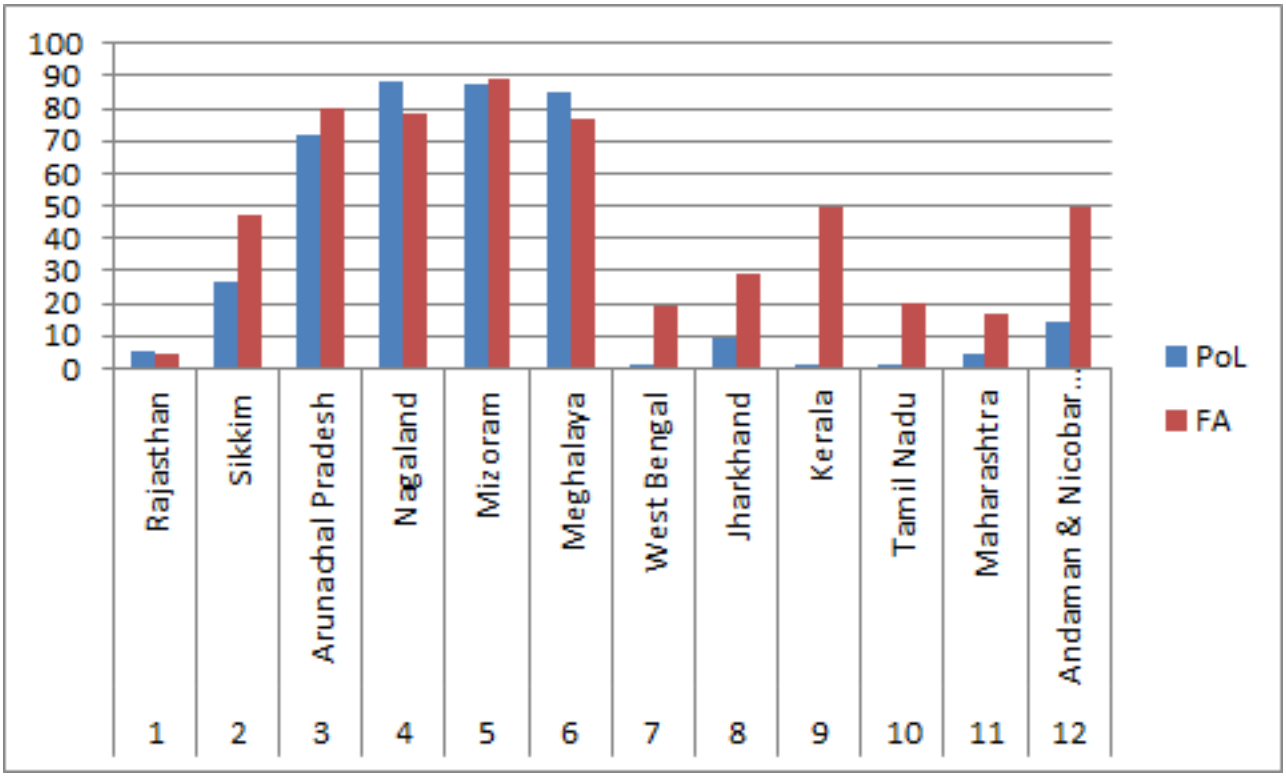
Sl. No.	State	PoL	FA
1	Rajasthan	5.36	4.73
2	Sikkim	26.36	47.31
3	Arunachal Pradesh	72.13	80.30
4	Nagaland	88.13	78.21
5	Mizoram	87.65	88.93
6	Meghalaya	85.35	76.76
7	West Bengal	0.53	18.96
8	Jharkhand	9.49	29.45
9	Kerala	0.43	49.50
10	Tamil Nadu	0.07	20.26
11	Maharashtra	4.57	16.45
12	Andaman & Nicobar Islands	14.77	49.50

Table 1: Data on PoL (population of other languages out of the total population of the Indian states and union territories according to the census 2011) and FA (forest area cover out of total area of the Indian states and union territories according to Compendium of Environment Statistics (2015) among 12 Indian states and union territories.

Sl. No.	State	PoL	Rural population growth rate (2011)
1	Rajasthan	5.36	18.95
2	Sikkim	26.36	-4.98
3	Arunachal Pradesh	72.13	22.52
4	Nagaland	88.13	-14.51
5	Mizoram	87.65	17.18
6	Meghalaya	85.35	27.13
7	West Bengal	0.53	7.67
8	Jharkhand	9.49	19.58
9	Kerala	0.43	-25.88
10	Tamil Nadu	0.07	6.60
11	Maharashtra	4.57	10.35
12	Andaman & Nicobar Islands	14.77	1.67

Table 2: Decadal growth rate of the rural population of other languages out of total population and rural population growth in the contemporary decade (2011) among the 12 Indian constitutional areas, i.e., state and union territories.

Figure


100
90
80
70
60
50
40
30
20
10
0
Rajasthan
Sikkim
Arunachal Pradesh
Nagaland
Mizoram
Meghalaya
West Bengal
Jharkhand
Kerala
Tamil Nadu
Maharashtra
Andaman & Nicobar...
1
2
3
4
5
6
7
8
9
10
11
12
PoL
FA