LANGUAGE DATA IN NATIONAL POPULATION AND HOUSING CENSUS OF NEPAL 2021: A CRITICAL ANALYSIS

Krishna Prasad Chalise

Central Department of Linguistics, Tribhuvan University
krishna.chalise@cdl.tu.edu.np

Analysis of NPHC 2021 language data based on retention of ancestral language as MT and MT in ethnic population, increase in MT and SL speakers after the previous census. Among them 39, 25, and 53 languages have problems with the first, second, and third variables respectively; and all have problems with the fourth variable. Unclear questions, improper technical word definitions, organized activities of caste/ethnic organizations and activists to influence the answers, and insufficient training for the enumerators are responsible for the problems. The post-enumeration survey did not include language data for validity and reliability evaluation. Future language policies and planning based solely on census data will lead to serious problems.

Keywords: NPHC 2021, EGIDS, census evaluation, overcount, language questions

1. Introduction

The Constitution of Nepal 2015 has guaranteed rights to equality to all citizens intending to make Nepal an inclusive and prosperous country. In this regard, it has acknowledged Nepal as a multilingual nation and designated all of the mother tongues as national languages. Unfortunately, a large number of minority languages are in various stages of endangerment. To uphold the spirit of the constitution, it is an imperative responsibility of the nation to protect and promote every mother tongue that is spoken in Nepal to ensure linguistic rights, linguistic mainstreaming, reduction of linguistic inequalities, and education in the mother tongue (the easiest language of the children) for quality education. The nation needs exhaustive and reliable data to design effective language policy and successful implementation to achieve the goals of protecting and promoting national languages. For these purposes, the National Population and Housing Census of Nepal 2021 sought to gather data on the ancestor's language, mother tongue, and second languages of the respondents.

However, the language data have always been fluctuating a lot across the censuses of Nepal in the past. These heavy fluctuations of language data in different censuses point out the severe validity and reliability crisis with the data. It was anticipated that the National Population and Housing Census of Nepal 2021 would supply sufficient and reliable data for national planning and policy implementation to achieve the goals outlined in the constitution.

The population census is a vital instrument for governments. It provides important details about geographic distribution and sociodemographic structure. In several instances, the census is the only data source for accurate estimations and social indicators. According to the UN (2017), housing and population highlights four reasons why housing and population counts are important. Firstly, it is crucial in public administration to ensure equity in the distribution of wealth, government services, and social inclusion. Secondly, all aspects of the national statistics system, including the social and economic ones, depend on it. Thirdly, it generates statistics on small areas and small population groups with no/minimum sampling errors. Fourthly, research and analysis use the census figures as authentic references. Therefore, it is imperative to guarantee the accuracy of the population and housing census data as substandard results might have catastrophic consequences for the entire planning and execution of national policies.

Any population and housing census is not perfect because of errors at different stages of census

operations. However, by thorough evaluation of the census results, we can identify the errors and their sources, and measure the errors to reduce their impacts on the planning and execution of national policies as much as possible. The census errors are broadly classified into coverage errors and content errors (UN, 2017). Coverage errors are related to omission or duplication in counting, and content errors are related to the deviation of information from reality or misinformation.

UN (2017) has recommended conducting a post-enumeration survey and demographic analysis for census evaluation to ascertain the quality of the census. The post-enumeration survey measures the accuracy of the census by independently surveying a sample of the population and matching each individual who is enumerated in the post-enumeration survey with information from the main enumeration to estimate the coverage and content errors in the census. Similarly, demographic analysis for census evaluation employs different demographic techniques ranging from visual inspection of census data to comparative analysis of the values in two censuses as well as demographic data from other sources. The National Population and Housing Census of Nepal 2021 conducted a post-enumeration survey (PES Nepal Report, 2023) and enumeration observation (Chalaune et al., 2022) to evaluate the quality of the National Population and Housing Census of Nepal 2021. The post-enumeration survey assessed the counting content errors regarding only age, sex, marital status, literacy status, and birthplace. It has not evaluated the errors in language-related issues. The enumeration observation reports the enumeration process of the Nepal National Population Census 2021, analyses the challenges faced in different stages of enumeration and adequacy of the census tools, and describes the activities of other groups and organizations that could affect the quality of the census. It hints at some activities that will affect the reliability and validity of the language-related data. However, neither evaluation measure could specify the quality of the language-related data.

2. Methodology

This research has assessed the content errors in the linguistic data in the National Population and Housing Census of Nepal 2021 based on the demographic analysis for the census evaluation method suggested by the UN (2017). It mainly employs the techniques of evaluating internal consistency within data and comparing the data in the new census with the corresponding data in previous censuses and the data from other reliable sources. I have evaluated the internal consistency within and between variables and compared the NPHC 2021 data with the corresponding data in the NPHC 2011 and the data from the sociolinguistic survey of the languages of Nepal carried out by the Linguistic Survey of Nepal, Central Department of Linguistics, Tribhuvan University (2008-2018), and Ethnologue: Languages of Nepal (Eppele et al., 2012).

This research has set four variables to assess the content errors in the census data. The first variable is the percentage of mother-tongue (MT henceforth) speakers among the respondents who accepted the language as the ancestor's language.

The second variable is the percentage of mother-tongue speakers in the total population. These variables indicate the portion of the mother tongue speakers among the ethnic population. The percentage of mother tongue retention in the total population can be calculated only if the population that a specific language speaks is specified. In this census, it can be calculated in the languages where there is a one-to-one correspondence between language and ethnicity. The reliability of the figures has been verified by comparing them with the vitality state of the corresponding languages identified by LinSuN (2008-2018) sociolinguistic reports and Eppele et al. (2012). Moreover, the similarity between the corresponding figures in the first and second variables determines the reliability of the data. This variable does not apply if the population of the group (caste/ethnicity) who speak the language is unavailable in the census.

The third variable is the increase in the number of mother tongue speakers in one decade (the difference between the 2021 and 2011 censuses).

The consistency of increments guarantees the data's dependability. Furthermore, the increments must correspond with the nation's average population growth rate. The population growth rate in the 2011 census was 1.35% per annum; in the 2021 census, it was 0.92%. The average population growth is 1.135% per annum and the total population growth in 10 years is 11.35% of the population. The population growth rate may vary by ethnicity and settlement, so the 8-15% increment can be the most reliable figure. The greater deviation of the increment rates suggests less reliability in the census data. Furthermore, there should be a reasonable correlation between the population and mother tongue growth percentages. This variable does not apply if the population of the group (caste/ethnicity) who speak the language is unavailable in the census.

The fourth variable is the increment in the number of second-language (SL henceforth) speakers. A decline in the number of MT speakers an increase in the SL speakers, and an unjustifiable increase in SL speakers suggest a lack of data reliability.

3. The analysis and findings

For analysis, all the languages enlisted in the census have been categorized into four categories based on the vitality state estimate of the languages of Nepal carried out by Eppele et al. (2012) based on the Expanded Graded Intergenerational Disruption Scale (EGIDS), a scale devised by Lewis and Simons (2010). This scale is an adaptation and expansion of the Graded Intergenerational Disruption Scale (GIDS) developed by Fishman (1991). Lewis and Simons (2017) proposed a new viewpoint named the Sustainable Use Model (SUM) and incorporated it with EGIDS for the evaluation of the vitality and development state of minority languages.

EGIDS levels 1 to 5 are included in the first category, level 6a, level 6b, and levels 7-9 are included in the second, third, and fourth categories respectively. The first group corresponds to the safe category, the second group to the unsafe category, the third group to the definitively endangered category, and the fourth group to the endangered (both severely endangered and critically endangered) group in the Language Vitality Assessment Scale proposed by UNESCO Ad Hoc Expert Group on Endangered Languages (Brenzinger et al., 2003).

The working definition of the second language made by NPHC 2021 is erroneous in itself. In reality, the language one is most fluent in after his/her mother tongue is the second language but the working definition made by NPHC 2021 is "Second language is the language spoken in the neighborhood or community except his/her mother tongue." It means the language may not be spoken by the respondents. For this reason, the data obtained as the second language speakers' data is not the second language speakers' data in itself. It can be assumed that this data may suggest what languages are spoken in a particular community. In the context of Nepal, it is most probable to be spoken more than two languages in a society. The definition allows recording only one language. For this reason, this data does not represent the languages spoken in a particular community. Because of the inherent problem in the working definition, the second language data is incongruent and it does not follow any logical pattern. Therefore, the fourth variable has not been analyzed for individual classes of languages.

3.1 EGIDS 1-5 languages

This group of languages, presented in Table 1, includes the EGIDS level 1 to 5 languages which more or less correspond to the 'Safe' category according to Brenzinger et al. (2003). According to Eppele et al. (2012), Nepali belongs to level 1 (National: the language that is used in education, work, mass media, and government at the nationwide level); Newar and Hindi belong to level 3 (Wider communication: the language used for local and regional work by both insiders and outsiders); Bhojpuri, Avadhi, Doteli, Urdu, Ranatharu, Bangla, Tibetan, and Munda belong to level 4 (Educational: the language being transmitted and standardization promoted through a system of institutionally supported education, situation of sustainable literacy); and Maithili[1],

[1] Maithili is taught at the school and university levels. It is surprising how Eppele et al. (2012) categorized it as an EGIDS level 5 language.

Tharu, Tamang, Rajbanshi, Sherpa, Thulung, Marwadi, and Tajpuriya belong to level 5 (Developing: the language used for face-to-face communication by all generations and has effective educational support in parts of the community). At present, Newar and Tamang have received the recognition of level 2 (Provincial: the language used for local and regional mass media and government services); and Maithili, Bhojpuri, Tharu, Awadhi, Doteli, Limbu, Khas, Bajjika, Gurung, and Magar have been recommended to be recognized as the provincial languages.

Table 1: *EGIDS 1-5 level languages*

MTs	% of MT		Increment % in 10 Years		
	In AL	In POP	MT	POP	SL
Nepali	129.07		10.63		55.28
Maithili	108.86		4.19		37.03
Bhojpuri	102.95		14.87		-13.13
Tharu	97.13	94.85	12.04	4.00	5.73
Tamang	88.78	86.77	5.15	6.49	113.95
Avadhi	94.78		151.53		66.52
Newar	73.17	64.36	1.98	1.46	0.03
Doteli	93.91		-37.18		
Urdu	72.64		-40.16		57.60
Rajbanshi	96.99	98.18	6.50	15.03	-79.54
Sherpa	91.75	90.24	2.67	15.66	15.32
Hindi	106.61		26.85		-81.80
Ranatharu	99.63				
Thulung	89.57	391.16	18.13	76.49	974.85
Bangla	89.59	172.27	12.88	-48.08	89.19
Marwadi	84.88	63.10		-34.29	
Tajpuriya	97.89	96.95	8.17	9.24	-82.21
Wambule	95.93		13.47		501.49
Tibetan	99.52		129.05		352.23
Koyee	93.19		226.67		361.69
Munda	90.27	58.70		52.72	
Lhomi	94.94	116.33			

Note. MT = Mother Tongue, AL = Ancestor's Language, POP = Population, SL = Second Language

3.1.1 Retention of ancestors' language as MT

In this group of languages, retaining ancestors' language as mother tongue is expected optimum. Regarding Nepali, Maithili, Bhojpuri, and Hindi, more people speak the languages as their mother tongue than those who said it was their ancestor's language. The languages are in wider use and spoken as the mother tongue by people from other ethnicities, too. On the contrary, Newar and Urdu have a low proportion. Newar population is widely distributed in the country, but the Newar language is very prominent only in the Kathmandu Valley and other dense Newar settlements. Similarly, the Muslims residing in the mountains of Nepal regard Urdu as their ancestors' language but it has not been in use in the community for several generations. Despite the low proportion, Newar and Urdu are safe languages. The other languages have an 85-99% retention to justify themselves as safe languages. Based on this variable, the census data is reliable in the case of these languages.

3.1.2 Speakers' percentage in ethnic population

In safe languages, the ethnic population is almost equal to the number of MT speakers if it is not used as MT by other ethnic groups. If the language is prominent in its native areas and weak in some marginal areas, the language is still safe although the number of MT speakers is slightly lower than the ethnic population. In this respect, the census data about Tharu, Tamang, Newar, Rajbanshi, Sherpa, and Tajpuriya is logical. The number of MT speakers is higher in Thulung, Bangla, and Lhomi. It was possible if the languages were used as MT by other ethnic groups, but the languages are not interethnic and also are not used as the languages of wider communication. The percentage of MT speakers in the ethnic population in Marwadi and Munda is extremely low to justify them as the safe languages. In addition, the retention of ancestors' language as MT and the percentage of MT speakers in the ethnic population in the languages do not match each other. Therefore, the data related to Munda and Marwadi is problematic.

3.1.3 Increase in MT speakers

Increment in the mother tongue speakers is normal with Nepali, Bhojpuri, Tharu, Bangla, and Wambule. It is not reliable with the other languages. It is negative with Doteli and Urdu, and extremely low with Newar and Sherpa. But it is relatively high with Thulung and Hindi, and impossibly high with Tibetan, Avadhi, and Koyee. If the proportion of mother tongue speakers within the total ethnic population, an

increase in mother tongue speakers, and an increase in population are compared, there is a mismatch with other languages, too.

3.2 EGIDS 6a languages

EGIDS 6a level (Vigorous) includes the languages used for face-to-face communication by all generations and being learned by children as their first language. This stage is called the situation of sustainable orality. The languages that belong to this group are presented in Table 2. This group of languages corresponds to the 'Stable yet threatened' category of languages according to Brenzinger et al. (2003).

Table 2: *EGIDS 6a level languages*

MTs	% of MT		Increment % in 10 Years		
	In AL	In POP	MT	POP	SL
Bajjika	98.93		33.92		41.40
Magahi	54.32		546.14		169.96
Baitadeli	37.18		-43.98		
Achhami	35.67		-0.94		
Khash	35.98		6626.44		6.93
Bajhangi	46.97		47.42		
Magar Kham	71.11		238.40		
Bajureli	46.98		427.70		
Santhali	94.80	93.66	7.65	10.77	-28.04
Darchuleli	29.61		670.05		
Ghale	79.63	65.04	184.83	54.86	234.37
Dadeldhuri	28.38		4264.75		
Dhimal	83.25	80.26	6.64	-2.49	-11.2
Khaling	90.87	280.42	14.14	274.8	788.60
Bahing	94.49	220.69	23.94	111.4	325.22
Hyolmo	92.50	98.36	-5.09	-8.67	5.39
Jumli	17.22		879.78		
Dailekhi	16.42		60.83		
Chum/Nubri	98.52	97.05			
Kewarat	98.88	39.38			
Dolpali	66.54	55.75	94.60	41.66	
Lhopa	98.48	168.92	-22.48	-47.0	
Sign Lg.	275.73		-60.14		2265.71
Byansi	77.89		255.41		-55.55
Baragunwa	100.91				
Sadri	120.91				
Magar Kaike	60.34		2350		2929.41
Kisan	86.47	67.88	-14.77	-14.9	-26.66
Dhuleli	105.50		110.16		
Raute	89.60	130.91	60.73	-8.41	1182.92
Lowa	96.29	44.89			
Kagate	112.52				
Nar-Phu	92.44				
Tichh. Poike	99.75				

3.2.1 Retention of ancestors' language as MT

The languages in this group are supposed to be spoken by all generations of people in the community. Therefore, the percentage of the mother tongue speakers should be optimum both in total ancestors' language figure and total population. However, in Magahi, Baitadeli, Achhami, Khash, Bajhangi, Bajureli, Darchuleli, Dadeldhuri, Jumli, and Dailekhi the percentage of mother tongue speakers is too low in total ancestors' language figure. According to Regmi (2021), all generations speak these languages without disrupting intergenerational language transmission. Chalise (2013) reports the same for Bajhangi and Bajureli in detail. Similarly, the percentage of Magar Kham, Magar Kaike, Ghale, Dolpali, and Dhimal is not high enough to justify their vigorous status. The ancestors' language retention in Nepali Sign Language is 275.73% is surprising. How can a sign language become an ancestors' language?

3.2.2 Speakers' percentage in ethnic population

The percentage of MT speakers in the total ethnic population of Khaling, Bahing, Lhopa, and Raute is impossibly high. The evidence from other studies justifies that these languages are unlikely to have many more speakers than the total ethnic population. Rai (2015) reports that Nepali is the only language of wider communication in the Khaling community. If the members of other communities spoke Khaling, Nepali would not have been the only language of wider communication in the community. The vitality of the language even in the Khaling community is not very strong as only 68% of the respondents can speak Khaling very well and nearly only 70% of the children can speak the language. Endangered Languages Project (2024) categorizes Bahing as a critically endangered language. If it had been spoken by the people twice the ethnic population, it would not have been a critically endangered language. Regmi et al. (2017) justify that Lhopa is a fully vital language, but they state that Nepali exclusively serves as the language of wider communication in the community. It means that Lhopa is spoken only by the Lhopa community. In this case, the number of speakers more than one and a half multiple of the ethnic population cannot be real data. It is impossible to speak Raute as the mother tongue by anyone out of the Raute community still, the number of mother tongue speakers is 30% greater than the

total Raute population. Bandhu et al. (2012) point out that Nepali and Doteli are used as the languages of wider communication and the Raute children are shifting to Nepali. The available facts suggest that the number of Raute mother tongue speakers in the census data is not reliable.

Similarly, in the languages, the percentage of MT speakers in the total ethnic population does not match the percentage of the speakers in the number who identified the languages as the ancestors' language. With Lhopa, the percentage of mother tongue speakers in the total ethnic population is only 44.89 but the percentage of retaining ancestors' language as mother tongue is 96.29%.

3.2.3 Increase in MT speakers

The increment in the mother tongue presents a dreadful picture of the census data. The increased percentage of mother tongue speakers ranges from -43.98 to 6626.44. If all the newly born children within one decade spoke the language as their mother tongue, the increase in mother tongue speakers is almost equal to the average population growth of the country within the decade. Similarly, if all the newly born children did not speak the language, the decrease in mother tongue speakers' percentage cannot be so large. Furthermore, the population growth percentage in ten years varies from -47.00 to 274.8. The correlations between the population growth and MT speakers' growth present a remarkable lack of correlation between them.

3.3 EGIDS 6b languages

EGIDS 6b level (Threatened) includes the languages that are used for face-to-face communication by all generations but only some of the children are learning it. The languages that belong to this group are presented in Table 3. This group of languages corresponds to the 'Unsafe' category of languages according to Brenzinger et al. (2003).

Table 3: *EGIDS 6b languages*

MTs	% of MT		Increment % in 10 Years		
	In AL	In POP	MT	POP	SL
Magar Dhut	64.92		2.76		26.05
Limbu	85.76	84.50	1.98	7.07	-11.24
Gurung	70.14	60.33	0.75	4.04	3.78
Bantawa	91.18	877.93	4.08	241.42	52.97
Chamling	88.37	731.12	15.93	82.63	360.02
Chepang	74.90	69.21	20.45	23.34	-26.60
Uranw/Urau	95.65	82.99	15.51		-64.54
Kulung	93.65	113.54	14.29	16.68	334.77
Angika	122.76		93.75		254.77
Majhi	52.19	29.56	34.78	32.99	-26.43
Sunuwar	73.87	41.44	-13.69	41.63	2.56
Thami	85.24	81.86	15.78	14.20	123.69
Sampang	82.72	446.12	18.21		612.69
Yakkha	78.57	81.56	-27.18	-28.25	-52.75
Darai	76.27	65.02	4.10	11.35	69.82
Yamphu	89.35	117.92	16.68	31.41	50.60
Bote	78.56	68.28	-12.30	8.28	937.6
Mewahang	90.43	129.70	59.74	84.74	1523.45
Puma	91.29		1.15		1247.31
Pahari	55.99	39.60	71.94	10.28	-38.26
Athpahariya	89.48	94.93	0.90	-1.65	
Dungmali	89.05		-13.69		508.13
Jirel	88.33	85.67	6.99	4.45	142.33
Chhantyal	56.88	35.79	-0.02	1.29	34.47
Raji	90.43	82.86	13.01	21.01	40.74
Meche	88.54	80.93	-3.93	6.69	-48.97
Lohorung	78.86	149.49	4.52	125.32	141.08
Jerung	93.68		250.80		-65.59
Chhintang	91.21		-30.92		111.80
Manange	94.92		415.81		
Walung	72.37	113.30	-53.37	-61.48	2663.63
Belhare	102.31		-70.45		12325
Khariya	60		-44.53		-100

3.3.1 Retention of ancestors' language as MT

In this group of languages, the number of MT speakers is expected to be less than the total ancestors' language figure as well as the total ethnic population because only some of the children can speak the language.

The percentage of MT speakers in the ancestors' language figure in Angika and Belhare exceeds 100%. Regmi & Chalise (2017) identifies that Angika is not a threatened language as suggested by Eppele et al. (2012). It is a vigorous language with sustainable orality. However, it is not clearly stated whether this language is spoken by people from other ethnic groups. If it was spoken by

people from other ethnic communities, the data could be justified. Endangered Languages Project (2024) categorizes Belhare as a severely endangered language. According to Regmi and Chalise (2017), 17% of the respondents do not speak their mother tongue every day and 57% of the children do not speak the language at all. They have shifted to Nepali, the language of wider communication. The evidence does not support the 102.31% of retention in Belhare. Similarly, the percentage in other languages like Bantawa, Urau, Kulung, Mewahang, Puma, Jerung, Chhintang, Raji, Manange, Athpahariya, Dungmali, etc. is more than or near to 90%. If we suppose that these figures are logical, we should not suppose that these languages are threatened ones.

3.3.2 Speakers' percentage in ethnic population

The percentage of MT speakers in the total ethnic population of Bantawa (877.93), Chamling (731.12), Kulung (113.54), Sampang (446.12), Yamphu (117.92), Mewahang (129.70), Lohorung (149.49), and Walung (113.30) is more than 100. It means that the languages are spoken by the population greater than the total ethnic population. It is possible if the languages are spoken by multi-ethnic groups.

Endangered Languages Project (2024) categorizes Bantawa, Kulung, Yamphu, Lohorung, and Walung as the threatened languages; Chamling as a vulnerable language; and Sampang as a severely endangered language.

Bantawa is reported to be spoken by other Rai communities, too, therefore there are probably more speakers of Bantawa than the total number of Bantawa people. But Rai et al. (2016) present a different scenario. According to them, only 82% of the respondents speak Bantawa every day, Nepali is the language of wider communication for all of them and 100% of them use Nepali every day. Moreover, only 63% of them are fully proficient in Bantawa, 17% of the children do not speak it, and 25% of young people do not speak it properly.

Rai (2017) reports that Chamling does not have a strong vitality state. Only 30% of the respondents have high proficiency in the language and 17% of them cannot speak the language at all. Similarly, 83% of the children do not speak the language and the parents speak Nepali with them. Nepali is the only language of wider communication.

Unlike Eppele et al. (2012), Yadav (2014) reports a vigorous status of Kulung. He states that Kulung people exclusively use Nepali as the language of wider communication. It suggests that the language is used only within the Kulung community.

Rai et al. (2015) report that more than 30% of children do not speak Sampang at all, and 20% of young people, too, do not speak it properly. They exclusively use Nepali with people from other language communities.

Hilty and Mitchell (2014) present different vitality states of Yamphu at various locations. However, on average, only 85% of the children can speak the language, 50% of the young people do not speak the language properly, and only 54% of the parents speak Yamphu with their children. They exclusively use Nepali with the members of other language communities.

Rai et al. (2014) report that 32% of the children do not speak Mewahang, 22% of young people do not speak it properly, 48% of the parents speak Nepali with their children, and Nepali is the only language of wider communication. Sharma (2022) remarks on the number of Mewahang speakers in NPHC (2011) that since there are no more than 500 proficient Mewahang speakers, the number of Mewahang speakers needs to be significantly revised.

Mitchell and Hilty (2014) present different vitality states of Lohorung in three different places. On average, they report that more than 50% of the children do not speak Lohorung at all, nearly 50% of the respondents assert that the young people do not speak it properly, and about 45 % of the parents do not speak it with their children.

Clark (2019) states that Walung is losing its vitality as it has been losing speakers in the youngest generation due to migration. S/he reports that varying in different locations, 11-40%

of the children do not speak it, 34% of the young speakers do not speak it properly, and 10% of the parents use Nepali with the children.

The information from Eppele et al. (2012), Endangered Languages Project (2024), the sociolinguistic survey reports on the individual languages by the Linguistic Survey of Nepal, and other sources presented here support each other and collectively they do not justify the higher number of speakers than the total ethnic population in the languages as reported in the census.

3.3.3 Increase in MT speakers

In 10 years, the increase in MT speakers and ethnic population and their correlation seems unnatural in several languages. The MT speakers increment in Angika, Mewahang, Pahari, Jerung, Majhi, and Manange, and the decrease in Yakkha, Chintang, Walung, Belhare, and Khariya are ridiculously high. Ethnic population increase in Bantawa, Chamling, Sunuwar, Yamphu, Mewahang, and Lonorung is unnaturally high, and in Chepang and Raji is fairly higher than expected. It is fairly lower than the average national population growth rate in Gurung, Jirel, Chhantyal, and Meche. Similarly, the population decrease in Yakkha and Walung cannot be justified. The increase in MT speakers and ethnic population in Bantawa, Chamling, Sunuwar, Yamphu, Mewahang, Pahari, and Lohorung do not correlate.

3.4 EGIDS 7-9 languages

This group of languages, presented in Table 4, includes the EGIDS level 7 to 9 languages. The level 7 (Shifting) languages are used by the child-bearing generation among themselves but it is not being transmitted to their children; the number of domains the language is used in is decreasing. Danuwar, Kumal, Bhujel, Nachhiring, Dumi, Thakali, Done, Mugali, Lapcha, Chiling, Sonaha, Hayu, Phangduwali, and Surel belong to this group. In level 8a (Moribund) languages the only remaining active users of the language are members of the grandparent generation and older. Tilung belongs to this group. In level 8b (Nearly Extinct) languages the only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language. Baram, Lungkhim, Sam, Bankariya, and Kusunda belong to this group. Similarly, the level 9 (Dormant) languages serve as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency. Sanskrit and Dura belong to this group.

Table 4: *EGIDS 7-9 level languages*

MTs	% of MT		Increment % in 10 Years		
	In AL	In POP	MT	POP	SL
Danuwar	81.65	60.38	9.10	-1.58	-70.30
Kumal	36.57	14.21	50.83	7.01	-56.29
Bhujel	34.20	10.88	-39.73	1.34	-53.51
Nachhiring	90.40	135.69	-1.34	2.04	336.86
Dumi	86.30		13.09		477.18
Thakali	48.18	35.94	-19.49	-11.15	-4.05
Done	88.09	145.88			
Mugali	97.25	133.42			
Lapcha	65.68	62.60	-70.12	3.86	3.86
Chhiling	75.06		-1.71		247.71
Sonaha	96.64		104.14		
Hayu/Vayu	45.46	36.91	-25.46	4.92	675.55
Phangduwali	102.48		-14.82		
Surel	86.56	54.71	-39.37		
Tilung	92.74		38.27		743.06
Baram	29.42	19.58	892.90	-3.45	458.18
Lungkhim	95.12		444.18		55.55
Sam	64.63		-73.56		690
Bankariya	104.87	47.77	24.63		
Kusunda	26.43	9.09	-17.85	-7.32	220
Sanskrit	10.26		733.19		122.35
Dura	48.02	35.67	-7.65	3.46	61.62

3.4.1 Retention of ancestors' language as MT

'The people speak shifting' languages up to the child-bearing generation. It means the young generation people do not speak the language. If the youngest child-bearing age is supposed to be 20 and the average life expectancy is 80 years, the portion of the young generation who does not speak the language is 20%. Therefore, language retention is expected to be a maximum of around 80% or less. Based on this parameter, the retention of ancestors' Language as MT in Nachhiring, Dumi, Done, Mugali, Phangduwali, Sonaha, and Surel seems to be overcounted. As a moribund language (only the grandparent generation speaks the language), the retention of Tilung also seems overcounted. In the nearly extinct languages Baram, Lungkhim, Bankariya, and Kusunda are also overcounted.

3.4.2 Speakers' percentage in the ethnic population

The percentage of MT speakers in the ethnic population in Nachhiring, Done, and Mugali exceeds 100% which is impossible. Likewise, the level of MT retention in Baram, Bankariya, and Dura is also not convincing.

3.4.3 Increase in MT speakers

Out of the 20 languages in this group, the increment in the number of MT speakers in 13 languages is not logical. The MT speakers increment in Kumal, Sonaha, Tilung, Baram, Lungkhim, and Sanskrit is naturally impossible. The comparison between the MT speaker increment and the ethnic population increment also justifies that it is unreliable. The increment in Bankariya is fairly higher than the average population growth and it is not spoken by any other communities. In this context, the MT speaker increment in Bankariya is also not logical. The decrease in the number of MT speakers in Bhujel, Lapcha, Hayu, Surel, and Sam is not logical and cannot be accepted. The comparison between the MT speaker increment and the ethnic population increment also justifies that it is unreliable.

4. Summary, discussion, and conclusions

Retaining ancestors' language as mother tongue in EGIDS 1-5 languages is logically convincing. In EGIDS 6a languages, it is highly undercounted in the languages previously regarded as the dialects of Nepali. Similarly, the retention percentage in Magar Kham, Ghale, Dolpali, and Magar Kaike is low to justify their vigorous status. In EGIDS 6b languages, ancestors' language retention is over-counted in Angika, Belhare. Mewahang, Puma, Athpahariya, Jerung, and Chhintang. In EGIDS 7 languages there is overcount in Nachhring, Dumi, Done, Mugali, Sonaha, Phangduwali, and Surel, Likewise, the data is overcounted in the EGIDS 8a language Tilung, 8b languages Baram, Lungkhim, Sam, Bankariya, Kusunda, and 9 languages Sanskrit and Dura.

MT retention in the ethnic population in EGIDS 1-5 level languages is overcounted in Thulung, Bangla, and Lhomi, and undercounted in Marwadi and Munda. In EGIDS 6a level languages, it is overcounted in Khaling, Bahing, Lhopa, and Raute, and undercounted in Lowa. In EGIDS 6b level languages, data is overcounted in Bantawa, Chamling, Kulung, Sampang, Yamphu, Mewahang, Lohorung, and Walung. In EGIDS 7-9 level languages, there is an overcount in Nachhiring, Done, Mugali, Phangduwali, Baram, Bankariya, and Dura.

Change in MT speakers' percentage in comparison with the 2011 census in EGIDS 1-5 level languages is overcounted in Avadhi, Tibetan, Koyee and undercounted in Newar, and negative in Doteli and Urdu. In EGIDS 6a level languages it is too problematic. It is overcounted in most of the languages like Magahi, Khas, Bajhangi, Magar Kham, Bajureli, Darchuleli, Ghale, Dadeldhuri, Bahing, Jumli, Dailekhi, Dolpali, Byansi, Magar Kaike, Dhuleli, and Raute. It is reduced in Baitadeli, Lhopa, and Sign language. In EGIDS 6b level languages it is overcounted in Angika, Majhi, Mewahang, Pahari, Jerung, Manange, and undercounted in Yakkha, Dungmali, Chhintang, Walung, Belhare, and Khariya. In EGIDS 7-9 level languages, the increase in Kumal, Sonaha, Tilung, Baram, Lungkhim, Bankariya, and Sanskrit and the decrease in Bhujel, Lepcha, Hayu, Phangduwali, Surel, Sam, Kusunda, and Dura is not convincing.

Changes in Second language speakers' percentage in comparison with the 2011 census in any of the languages have no logical justification.

From the analysis, it can be clearly understood that there are several problems in the census report of Nepal 2021 regarding the language data. There are several factors responsible for the deviation of the data. The first responsible factor is the problems with the set of language-related questions. The statement made by Chalaune et al. (2022) that the enumerators were especially confused about the questions regarding emigration, economic activities, child mortality, language, etc.

The first question was "What is your ancestral language?" It is problematic to specify the ancestors. It is believed that the ancestors of the

people speaking the languages of a language family were common. Is it logical to state that the ancestral language of Indo-Aryan language-speaking people is Sanskrit? Another serious problem is regarding the ancestral language of the speakers of Bajhangi, Doteli, Jumli, etc. Previously the languages were identified as the dialects of Nepali therefore their parents, grandparents, and so on spoke Nepali. In this sense, for Bajhangi speakers, it is their MT and Nepali is their ancestral language. But the fact is that the same language has been spoken for several generations. This is why the retention of ancestral language as the mother tongue is undercounted in the languages previously regarded as the dialects of Nepali. There are some communities not speaking the ethnic language for several generations. It is difficult to determine whether their ethnic language or the language they have been speaking for generations is their ancestral language.

The second question is "What is your mother tongue?" Some people perceive it as an ethnic language or language of identity no matter whether they speak it or not. Some perceive it as the language of their mother, and some perceive it as their most proficient language, and so on. The working definition made by the census is "the language spoken as the first language is the mother tongue. The working definition is confusing in itself. It is not clear whether it is the first language the respondent spoke in his/her life or the language the respondent is the most proficient in. However, in reality, the respondents perceived it as the ethnic language or the language of identity, and the number of MT speakers is overcounted in most of the minority languages. The term "mother tongue" has three different values: identity value, emotional value, and functional value. The census data regarding the mother tongue carries the identity value and/or emotional value. The working definition made by the NPHC tries to obtain the functional value which is further justified by the statement that the data will help to make policies and plans to assist promotion and preservation of the languages (Bhattarai et al., 2021)). For the functional value, the measurement of language proficiency is essential which is lacking. Because of the overcount in most of the minority languages and lacking in the measurement of language proficiency, the data loses its functional relevance.

The third question is "What is your second language?" The language one is most fluent in after his/her mother tongue is the second language but the working definition made by NPHC is "Second language is the language spoken in the neighborhood or community except his/her mother tongue." This definition is false in itself and, for this reason, the data obtained as the second language data is not the second language data in itself.

Undercount and overcount in the census data can also be attributed to the census campaigns by various groups of people. According to Chalaune et al. (2022), these groups distributed pamphlets appealing to the members of their communities about what to record as their ethnicity, language, and religion. They also conveyed the message through social media. In the case of language, they appealed to the people that they should record their ethnic or community's language as both the ancestors' language and mother tongue, and the language of neighboring Indigenous ethnicities as the second language. In addition to the campaigns, in some areas, different ethnic activists exerted pressure on enumerators to record particular responses to questions about language and religion. As the enumerators and supervisors were recruited from the respective communities as far as possible (Bhattarai et al., 2021), how much they could be neutral and resist the pressure from the ethnic organizations and activists is an important question.

The language questions were complex for the enumerators to understand. The training was planned for six days in reality it was conducted only for four days. According to Chalaune et al. (2022), some enumerators complained that the training was just a formality and even the census officials acknowledged that the training did not provide adequate practical exercise. It justifies that the enumerators were not trained enough to handle the language-related questions.

In my own experience, the enumerators did not visit my house. It was reported that they took information about my family from a nearby grocer. It is only a representative case and I have met several people with the same story in Kathmandu. PES Nepal Report (2023) shows that there is an undercount of data in the urban areas. The undercount in the Newar data seems to be because of this reason.

Some of the errors may have been routed from the previous censuses.

In the NPHC 2021, the linguistic data has several anomalies in addition to discrepancies between and among its different variables. If the country bases all future language policy and planning decisions only on census data, it would undoubtedly lead to serious problems in the future. UN (2017) suggests conducting a post-enumeration survey to ascertain the quality of the census data but the post-enumeration survey for NPHC 2021 did not include the language and several other issues to ascertain the quality of the data. In this situation, the safe way is to evaluate the validity of the census data comparing it with the data from other reliable sources for future language policy and planning decisions.

Future censuses should pay greater attention to choosing the questions and providing workable definitions for specific words. It is unethical to attempt to manipulate census data this or that way therefore the Census Bureau should prepare a measure to control it. Similarly, the census bureau is expected to improve the training for the enumerators and overall management of census data collection.

References

Bandhu, C. M., Sharma Pathak, L. N., & Rai, T. N. (2012). *A sociolinguistic survey of Raute:A Tibeto-Burman language* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Bhattarai, I., Chalaune, A., Subedi, S., & Limbu, A. (2021). *Nepal population and housing census 2021: Preparation and contentious*. Democracy Resource Center Nepal.

Brenzinger, M., Dwyer A., Graaf, T., Grinevald, C., Krauss, M., Miyaoka, O., Ostler, N., Sakiyama, O., Villalon, M., Yamamoto, A., & Zepeda, O. (2003). *Language vitality and endangerment. UNESCO Ad Hoc Expert Group Meeting on Endangered Languages*. http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf.

Chalaune, A., Ajeet, A. Limbu, A., Bhattarai, I. Manandhar, C., & Karki, J. (2022). *Nepal national population and housing census 2021: Enumeration observation.* Democracy Resource Center Nepal.

Chalise, K. P. (2013). *A sociolinguistic survey of Bajhangi and Bajureli* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Clark, M. (2019). *A sociolinguistic study of Walunge and related varieties Dhokpya and Thudam* [Unpublished research report]. Central Department of Linguistics Tribhuvan University, Nepal and SIL International.

Endangered Languages Project. (2024). *Catalogue of endangered languages*. University of Hawaii at Manoa. http://www.endangeredlanguages.com

Eppele, J. W., Lewis, M. P., Regmi, D. R., & Yadava, Y. P. (Eds.). (2012). *Ethnologue: Languages of Nepal*. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics and SIL International.

Fishman, J. A. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Multilingual Matters.

Hilty, H. J., & Mitchell, J. R. (2014). *Yamphu: A sociolinguistic survey*. SIL International. https://www.sil.org/system/files/reapdata/39/87/23/39872354340493988124485710717937538249/silesr2014_007.pdf

Language Commission, Nepal (2075 VS.). *History of Done language* [Unpublished report]. Language Commission, Sankhamul, Kathmandu.

Lewis, P., & Simons, G. (2010). Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique, 55*, 103-20.

Lewis, P., & Simons, G. 2017. *Sustaining language use: Perspectives on community-based language development.* SIL.

Mitchell, J. R., & Hilty, H. J. (2014). *Sociolinguistic survey of Lohorung*. SIL International.

National Statistics Office/NSO (2021). *National population and housing census-2021*. Retrieved from https://censusnepal.cbs.gov.np/.

PES Nepal Report. (2023). *Census of Nepal 2021: Report on post enumeration survey (PES)*. National Statistics Office.

Rai, N. K., Rai, N. M., & Thokar, R. (2014). *A sociolinguistic survey of the Mewahang language* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Rai, N. K., Rai, N. M., & Thokar, R. (2015). *A sociolinguistic survey of Sampang* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Rai, N. K., Rai, N. M., & Thokar, R. (2016). *A sociolinguistic survey of the Bantawa language* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Rai, T. M. (2015). *A sociolinguistic survey of Khaling: A Tibeto-Burman language* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Rai, T. M. (2017). *A sociolinguistic survey of Chamling: A Tibeto-Burman language* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Regmi, A., & Chalise, K. P. (2017) *A sociolinguistic survey of Belhare: A Tibeto-Burman language* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Regmi, D. R. (2021). *A sociolinguistic survey of the languages of Nepal: A synopsis, Vol. II. Indo-Aryan, Dravidian, and Austro-Asiatic languages*. Lincom GmbH.

Regmi, D. R., Gautam, B. L., & Regmi, A. (2017). *A sociolinguistic survey of Lhowa* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

Sharma, N. P. (2022). Comprehensive documentation of Mewahang: A first linguistic field report. *Nepalese Linguistics, 35,* 116-127.

UN. (2017). *Principles and recommendations for population and housing census, Revision 3.* UN, Department of Economic and Social Affairs: Statistics Division.

Yadav, S. P. (2014). *A sociolinguistic survey of Kulung* [Unpublished research report]. Linguistic Survey of Nepal (LinSuN), Central Department of Linguistics, T.U.

History of article

Received: July 01, 2024
Revised: October 15, 2024
Accepted: 05 November, 2024