Language Use and Choice Among the Semai Community in Kampung Sungai Perah, Parit, Perak

Noor Saadah Salleh[1*], Siti Nur Zulaika Binti Mohd Isyam[2]

Department of Malay Studies, Academy of Language Studies, Universiti Teknologi MARA, Shah Alam, Malaysia

*Corresponding Author

ABSTRACT

This study examines the phenomenon of language choice among the Semai Indigenous community in Kampung Sungai Perah, Parit, Perak, focusing on language proficiency and language use across different domains. Fishman's (1972) domain theory was employed as the conceptual framework to understand these dynamics. Data were collected through structured questionnaires from 112 respondents, representing 20% of the population, and analysed descriptively using SPSS. The findings indicate that Malay language is more dominant across most communication domains, while the Semai language remains in use within family settings and traditional ceremonies. These results highlight that the Semai language is endangered, particularly among younger generations, and underscore the need for integrated revitalization strategies to ensure the preservation of the community's language and cultural heritage.

Keywords: Language choice, Semai Indigenous community, Minority language, Language sustainability, Fishman's Domain Theory

INTRODUCTION

The Semai language is the mother tongue of the Semai community, which belongs to the Austroasiatic language family, specifically the Mon-Khmer branch. With 62,440 speakers, Semai is one of the largest Indigenous languages in Peninsular Malaysia (JAKOA,2024). As an oral language, it plays an essential role in preserving ethnic identity and serves as the primary medium of daily communication, particularly within the domains of family, community activities, and traditional rituals. The language functions not only as a tool of communication but also as a cultural heritage symbol that connects the Semai people to their ancestral roots.

Nevertheless, the position of the Semai language has become increasingly challenged due to the dominance of Malay, which functions as the national language, the language of education, and the language of social mobility. Recent research indicates that although Semai children still acquire Semai as their first language, proficiency in Malay is steadily increasing because of schooling, socialisation, and interethnic interactions (Sandai & Mahmud, 2023). This situation reflects the presence of unbalanced bilingualism, where Semai remains resilient in intimate domains such as family and rituals, yet is increasingly marginalised in official and public domains.

In this context, the phenomenon of language choice becomes a critical aspect to examine, as it reflects which language is prioritised situations. Semai language is often used to maintain ethnic solidarity and intragroup relationships, while Malay language is selected for education, employment, and broader social interactions. Thus, language choice is not merely a communicative decision but also reflects the broader challenges of sustaining minority languages within the forces of globalisation and social mobility.

Accordingly, this study focuses on language choice within the Semai Indigenous community in Kampung Sungai Perah Parit, Perak, employing Fishman's (1972) domain approach as the analytical framework. Through this perspective, patterns of language choice can be identified to assess the extent to which Semai is retained in daily

life and to evaluate the potential for language shift in the long term.

LITERATURE REVIEW

Studies on language choice among the Orang Asli community, particularly the Semai ethnic group, have highlighted the various factors that influence the use of minority languages. The Semai language functions as the mother tongue and the main medium in the domains of family, culture, and ritual, yet its position is increasingly challenged due to the influence of Malay. Sandai and Mahmud (2023) found that although Semai children still acquire Semai as their first language, proficiency in Malay is steadily increasing because of formal education and social interaction. This illustrates the presence of unbalanced bilingualism among the younger generation.

In addition, research conducted by Zailani, Makhtar, and Yusop (2022) on the Semai community in Hulu Selangor revealed that the Semai language remains dominant in intimate domains such as family and ritual ceremonies, but Malay language is preferred in education, employment, and inter-ethnic interactions. This pattern is consistent with Fishman's (1972) domain framework, which explains that majority languages tend to dominate strategic domains, while minority languages persist in private spheres. This situation suggests that the social status and functional role of a language determine the sustainability of Semai usage.

Furthermore, the study by Khairul Ashraaf Saari and Harishon Radzi (2023) examined language choice among Orang Asli participants in missionary training programs in Kok Lanas across three main domains: teaching and learning, inter-ethnic interaction, and social activities. Using a mixed-methods approach that combined questionnaires, interviews, and observations, the findings demonstrated that Standard Malay was chosen as the most effective language of interaction for enhancing participants' understanding, while Temiar and the Kelantan dialect continued to be used in specific contexts. These results underscore that language use is dependent on the underlying social context of each domain.

Meanwhile, in a broader context, the study by Muhammad Imran Afzal, Liaqat Ali Mohsin, and Sadia Asif (2022) on Punjabi speakers in Pakistan found that language shift occurs due to the dominance of majority languages. Although Punjabi is still spoken, many speakers are now regarded as semi-speakers because they increasingly shift to Urdu, English, and Arabic in particular domains. Domain analysis showed that the social status and functional role of languages play a crucial role in shaping language choice. This study demonstrates that the vitality of minority languages becomes weakened when they fail to maintain a foothold in strategic domains.

Overall, this review highlights a consistent pattern in which minority languages such as Semai function mainly in intimate domains like family and ritual but are increasingly marginalized in official and public spheres. Therefore, this study makes an important contribution by examining language choice among the Semai Orang Asli community in Kampung Sungai Perah Parit, Perak, using Fishman's (1972) domain framework to assess patterns of bilingualism and the potential for language shift in the contemporary context.

METHODOLOGY

This study employed a field-based mixed-methods approach that combined quantitative and qualitative techniques to provide a comprehensive understanding of language proficiency and language choice among the Semai Indigenous community in Kampung Sungai Perah, Parit, Perak. The use of mixed methods is considered appropriate as it allows for triangulation and provides a more holistic analysis of sociolinguistic phenomena (Creswell & Plano Clark, 2018; Tashakkori & Teddlie, 2010). A total of 112 respondents, representing 20% of the village population (561 individuals), were selected using simple random sampling to ensure equal chances of selection for all members of the population, a method widely recognized for reducing sampling bias (Etikan & Bala, 2017).

The main research instrument was a structured questionnaire consisting of three sections, namely demographic information, language proficiency, and language choice by domain. Structured questionnaires are effective tools in sociolinguistic research for capturing systematic patterns of language use across domains (Milroy & Gordon, 2003). Complementary qualitative data were obtained through field notes, with a focus on observing language

use in daily life, consistent with ethnographic approaches in language studies (Heller, 2008). All questionnaire data were digitized, screened for completeness, and analysed using SPSS with descriptive statistical methods, while the qualitative data were used to contextualize the quantitative findings. To ensure validity, the questionnaire was reviewed by experts in minority language studies, aligning with best practices in instrument development (Bryman, 2016)

RESULTS AND DISCUSSION

The Language Proficiency of the Semai Indigenous Community: Evidence from Kampung Sungai Perah, Parit, Perak.

Table 1 Language Proficiency of the Semai Indigenous Community

Language	1 (VLP)	2 (LP)	3 (MP)	4 (P)	5 (VP)	Mean
Semai Language (SL)	– (0%)	– (0%)	– (0%)	15 (13.4%)	97 (86.6%)	4.90
Malay Language (ML)	– (0%)	3 (2.7%)	16 (14.3%)	61 (54.5%)	32 (28.6%)	4.10
English Language (EL)	31 (27.7%)	22 (19.6%)	50 (44.6%)	9 (8.0%)	– (0%)	2.33
Perak Dialect (PD)	7 (6.3%)	42 (37.5%)	45 (40.2%)	12 (10.7%)	6 (5.4%)	2.71

Note. VNP = Very Not Proficient, NP = Not Proficient, LP = Less Proficient, P = Proficient, VP = Very Proficient

Table 1 shows that the Semai Language (SL) has the highest proficiency level (Mean=4.90), with almost all respondents rated as 'very proficient'. This indicates that the mother tongue remains strong within the community, particularly in the domains of family and culture. However, the Malay Language (ML) with a similarly high mean score (4.10) demonstrates that the Semai generation not only retains their mother tongue but also adapts to the majority language. This phenomenon highlights functional bilingualism, where the Semai language dominates the identity domain, while ML dominates formal domains.

In contrast, English Language (EL) (Mean=2.33) and Perak Dialect (PD) (Mean=2.71) show low proficiency levels, indicating they are not considered important for daily interaction. The low English proficiency reflects that the Semai community emphasizes the utilitarian function of ML over EL. From Fishman's (1972) perspective, these findings reveal that the mother tongue still functions in intimate domains, but the dominance of ML as the language of social mobility highlights a potential generational language shift.

Language Choice by Domain

Family Domain

Table 2 Language Choice in the Family Domain

Situation/Context	SL n (%)	ML n (%)	SL+ML n (%)	Total	%
Talking with grandparents	101 (90.18)	1 (0.89)	10 (8.93)	112	100
Talking with parents	102 (91.07)	– (–)	10 (8.93)	112	100
Talking with siblings	92 (82.14)	– (–)	20 (17.86)	112	100
Talking with spouse	44 (83.02)	2 (3.77)	7 (13.21)	53	100
Talking with close relatives	93 (83.04)	3 (2.68)	16 (14.29)	112	100
Talking with children	42 (80.77)	2 (3.85)	8 (15.38)	52	100

Note. SL = Semai Language; ML = Malay Language; SL+ML = Code-mixing of Semai and Malay.

The examination of the family domain is essential as it represents the primary site for intergenerational transmission of language and thus plays a critical role in the maintenance of minority languages such as the

Semai Language (SL). Situations with grandparents and parents demonstrate the dominance of SL (over 90%). This reflects the continuity of the mother tongue in communication with the older generation, indicating that the family domain remains a stronghold for language maintenance. The use of ML is almost non-existent, suggesting that the older generation functions as guardians of the Semai language.

However, communication with siblings, spouses, and children shows a slight decline in the use of SL (80–83%), with an increase in code-mixing between SL and ML. This indicates a generational shift, as ML begins to penetrate intrafamily communication despite SL remaining dominant. Such a pattern reflects the common situation where minority languages continue to be used at home but are increasingly mixed with majority languages among younger generations.

Similarly, interaction with close relatives (83% SL, 14% SL+ML) demonstrates the same trend. SL is still the primary choice, but ML increasingly appears in mixed forms, proving that ML is no longer confined to formal domains but is starting to influence wider family interactions.

In sum, the analysis demonstrates that while SL continues to maintain vitality within the family domain, signs of unbalanced bilingualism are increasingly evident as ML penetrates intergenerational communication. From Fishman's (1972) perspective, the family domain constitutes the cornerstone for the maintenance of minority languages; however, within the Semai community this function appears to be gradually weakening. Should this trend persist, it is plausible that future generations may come to prioritise ML for daily interactions, relegating SL to a more symbolic role despite continued comprehension.

Friendship Domain

Table 3 Language Choice in the Friendship Domain

Situation/Context	SL n (%)	ML n (%)	SL+ML n (%)	Total (N)	%
Talking with Semai friends	101 (90.18)	– (–)	11 (9.82)	112	100
Talking with non-Semai friends	11 (9.82)	94 (83.93)	7 (6.25)	112	100

Note. SL = Semai Language; ML = Malay Language; SL+ML = Code-mixing between Semai and Malay.

The friendship domain plays a crucial role in understanding language dynamics as peer interactions often shape language preferences among younger generations. Table 3 shows that language choice among peers is strongly influenced by the ethnic background of interlocutors. In communication with fellow Semai peers, SL is clearly dominant (90.18%) with minimal code-mixing (9.82%). This finding affirms that the mother tongue continues to function as a symbol of intimacy and solidarity in intragroup friendships.

In contrast, interactions with non-Semai peers show a starkly different pattern: ML is used extensively (83.93%), with SL limited to only 9.82% and code-mixing at 6.25%. This demonstrates that ML functions as the lingua franca in interethnic relationships, consistent with its role as the national and educational language.

In summary, the friendship domain reflects situational bilingualism: SL dominates intragroup interaction, while ML is prioritised in intergroup contexts. From Fishman's (1972) perspective, this shows that the friendship domain does not fully protect SL but instead opens pathways for ML dominance. In the long run, this pattern could accelerate language shift if interethnic interactions become more frequent in the community's daily life.

Neighbourhood Domain

Table 4 Language Choice in the Neighbourhood Domain

Situation/Context	SL n (%)	ML n (%)	SL+ML n (%)	SL+EL n (%)	Total (N)	%
Socialising with Semai neighbours	103 (91.96)	1 (0.89)	8 (7.14)	– (–)	112	100

Socialising with non-Semai neighbours	15 (13.39)	87 (77.68)	8 (7.14)	2 (2.24)	112	100
Talking with elderly neighbours	95 (84.82)	4 (3.57)	13 (11.61)	– (–)	112	100
Talking with young neighbours	93 (83.04)	7 (6.25)	12 (10.71)	– (–)	112	100

Note. SL = Semai Language; ML = Malay Language; EL = English Language; SL+ML = Semai-Malay code-mixing; SL+EL = Semai-English code-mixing.

Table 4 reveals that neighbourhood interactions continue to preserve the dominance of SL, particularly when communicating with Semai neighbours. A total of 91.96% of respondents preferred SL in this context, with only 7.14% engaging in code-mixing with ML. This suggests that SL serves as the language of solidarity within a homogenous community setting.

However, interactions with non-Semai neighbours shift dramatically towards ML dominance (77.68%), followed by SL (13.39%), SL+ML (7.14%), and SL+EL (2.24%). This reflects the role of ML as the lingua franca in multicultural neighbourhood contexts. Although SL+EL usage is minimal, it signals the presence of English influence in modern interactions, albeit marginal.

Age also appears to influence language use. While SL remains dominant with both elderly (84.82%) and young neighbours (83.04%), the use of ML and code-mixing is slightly higher among younger neighbours (6.25% ML; 10.71% SL+ML). This indicates that younger generations are more open to incorporating ML in daily communication.

In general, the neighbourhood domain reflects situational bilingualism. SL retains its role as a marker of ethnic identity, but ML dominates interactions with non-Semai neighbours and is gradually increasing among younger generations. According to Fishman (1972), the neighbourhood should act as a space for mother tongue maintenance, but these findings suggest that ML is increasingly reshaping linguistic balance in social spaces. If this trend continues, SL risks being marginalised in future neighbourhood interactions

Education Domain

Table 5 Language Choice in the Education Domain

Situation/Context	SL n (%)	ML n (%)	SL+ML n (%)	Total (N)	%
Interaction with teachers/lecturers	– (–)	109 (97.32)	3 (2.68)	112	100
Sharing ideas with Semai peers	49 (43.75)	57 (50.89)	6 (5.36)	112	100
Academic discussions with non-Semai peers	1 (0.89)	107 (95.54)	4 (3.57)	112	100
Classroom activities	3 (2.68)	101 (90.18)	8 (7.14)	112	100
Extracurricular activities	9 (8.04)	95 (84.82)	8 (7.14)	112	100

Note. SL = Semai Language; ML = Malay Language; SL+ML = Semai-Malay code-mixing.

Table 5 demonstrates the overwhelming dominance of ML in all educational settings. In formal interactions with teachers/lecturers, nearly all respondents used ML (97.32%), with only minimal code-mixing (2.68%). This affirms the status of ML as the formal academic language, irreplaceable by SL.

When interacting with fellow Semai students, SL is used moderately (43.75%), but ML remains slightly higher (50.89%). This suggests that even among peers sharing the same mother tongue, ML is prioritised due to the academic environment's demand for formality. In academic discussions with non-Semai students, ML becomes almost universal (95.54%), functioning as the undisputed lingua franca.

Both classroom and extracurricular activities reflect the same pattern: ML remains dominant (90.18% in class; 84.82% outside class), while SL use is very limited. Code-mixing appears at a low level (7–8%), signalling a gradual shift in linguistic identity among the younger generation.

Findings in the education domain highlight that ML has almost completely replaced SL as a medium of academic communication. From Fishman's (1972) perspective, the role of education in sustaining mother tongues is weak, as institutional pressures privilege majority languages. If this trend persists, SL may remain confined to informal settings, shrinking its functional space among young Semai speakers.

Employment Domain

Table 6 Language Choice in the Employment Domain

Situation/Context	SL n (%)	ML n (%)	SL+ML n (%)	Total (N)	%
Interaction with non-Semai co-workers	8 (13.79)	50 (86.21)	– (–)	58	100
Work communication with Semai co-workers	39 (67.24)	16 (27.59)	3 (5.17)	58	100
Communication with employers	15 (25.86)	43 (74.14)	– (–)	58	100
Casual conversations at work	27 (46.55)	29 (50.00)	2 (3.45)	58	100

Note. SL = Semai Language; ML = Malay Language; SL+ML = Semai-Malay code-mixing.

Table 6 shows that ML is clearly dominant in most workplace contexts, particularly in interactions with non-Semai colleagues (86.21%) and employers (74.14%). This underscores ML's role as the language of social mobility and formal communication in employment settings.

However, in communication with Semai co-workers, SL still plays an important role (67.24%), although ML is also present at a moderate level (27.59%), with some code-mixing (5.17%). This indicates functional bilingualism, where language choice depends on the social background of interlocutors.

In casual conversations at work, ML (50%) and SL (46.55%) are used almost equally, with minimal code-mixing (3.45%). This suggests that while SL retains significance in informal contexts, ML continues to consolidate its position.

Overall, the employment domain reinforces ML as the main language in formal affairs, while SL maintains its role in intragroup and informal interactions. From Fishman's (1972) framework, this pattern shows that the workplace not only strengthens ML as a prestige language but also indirectly limits the functions of SL. If the trend continues, SL risks being confined to "informal" usage within the community, lacking strength in formal domains.

Religious Domain

Table 7 Language Choice in the Religious Domain

Situation/Context	SL n (%)	ML n (%)	SL+ML n (%)	Total (N)	%
Reciting prayers in places of worship	83 (74.11)	21 (18.75)	8 (7.14)	112	100
Language used in sermons	21 (18.75)	86 (76.79)	5 (4.46)	112	100

Communication with non-Semai religious leaders	8 (7.14)	99 (88.39)	5 (4.46)	112	100
Communication with Semai religious leaders	83 (74.11)	20 (17.86)	9 (8.04)	112	100

Note. SL = Semai Language; ML = Malay Language; SL+ML = Semai-Malay code-mixing.

Table 7 shows clear distinctions in language use depending on roles and participants in religious events. In prayer recitations at places of worship, SL is dominant (74.11%), reflecting the role of the mother tongue as a vehicle of ritual and spiritual identity. However, in sermons, ML is dominant (76.79%), suggesting that the majority language is deemed more appropriate for formal knowledge dissemination.

This difference is even more evident in communication with religious leaders. When non-Semai leaders are involved, ML dominates nearly all interactions (88.39%), with SL used minimally (7.14%). Conversely, when the leaders are of Semai origin, SL regains dominance (74.11%), although ML remains present at a moderate level (17.86%). This indicates that linguistic identity is preserved when participants share the same cultural background.

Overall, the religious domain highlights a functional division: SL is used for ritual and ethnic identity, while ML dominates sermons and communication with external religious leaders. From Fishman's (1972) perspective, this shows that although SL remains relevant in religious settings, its usage increasingly depends on the cultural background of participants. If this trend continues, SL may become restricted to cultural symbolism without practical strength in the transmission of religious knowledge.

CONCLUSION

Overall, this study found that the Semai Indigenous community in Kampung Sungai Perah, Parit, continues to maintain a high level of proficiency in their mother tongue, particularly the Semai language (SL), which recorded near-universal competence among respondents. However, analysis of language choice across domains revealed an unbalanced bilingual pattern, whereby SL remains dominant in the domains of family, ritual, and intragroup interaction, while Malay language (BM) is increasingly dominant in the domains of education, employment, and interethnic relations. This phenomenon is consistent with Fishman's (1972) framework, which emphasizes that domains play a crucial role in determining the sustainability of minority languages. Although SL remains strong as a language of identity and ethnic solidarity, the penetration of BM into strategic domains signals the potential for generational language shift, especially when the majority language is perceived as more valuable for social mobility and educational advancement. In conclusion, the findings of this study underscore that the sustainability of SL depends on efforts to strengthen its presence across multiple domains, rather than being confined to domestic and ritual spaces, so that it does not merely serve as a cultural symbol but continues to thrive as a medium of daily communication for future generations.

REFERENCES

1. Creswell, J. W. (2014). Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.). SAGE Publications.
2. Bryman, A. (2016). Social research methods (5th ed.). Oxford University Press.
3. Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE Publications.
4. Devamany S. Krishnasamy & Asan Ali Golam Hassan. (2016). Masyarakat Orang Asli: Impak Program Penempatan Semula. UUM Press
5. David, M. K., & Dealwis, C. (2008). Language use in the domain of neighbourhood: The case of the Bidayuh in Sarawak. SEAMEO Regional Language Centre Journal, 39(2), 195–211.
6. Eberhard, D. M., Simons, G. F., &Fennig, C. D. (2020.). Ethnologue: Languages of the World (23rd ed.). SIL International
7. Etikan, I., & Bala, K. (2017). Sampling and sampling methods. Biometrics & Biostatistics International Journal, 5(6), 215–217. https://doi.org/10.15406/bbij.2017.05.00149
8. Fasold, R. (1984). The Socioliguistics of Society. Oxford. Blackwell.

9. Fishman, J. A. (1972). The sociology of language: An interdisciplinary social science approach to language in society. Newbury House.

10. Heller, M. (2008). Doing ethnography. In L. Wei & M. G. Moyer (Eds.), The Blackwell guide to research methods in bilingualism and multilingualism (pp. 249–262). Blackwell.

11. Jabatan Kemajuan Orang Asli (JAKOA). (2024). Taburan etnik Orang Asli mengikut etnik / sub-etnik mengikut negeri. Retrieved September 20, 2025, from https://www.jakoa.gov.my/umum/taburan-etnik-orang-asli-mengikut-etnik-sub-etnik-mengikut-negeri/

12. Khairul Ashraaf Saari & Harishon Radzi. (2023). Pemilihan Bahasa dalam Kalangan Peserta Latihan dan Dakwah OrangAsli, Kok Lanas Berdasarkan Tiga Domain. Jurnal Linguistik. Vol 27 (1) Mei 2023 (129-139)

13. Milroy, L., & Gordon, M. (2003). Sociolinguistics: Method and interpretation. Blackwell.

14. Noriah Mohamed. (2008). Language choice and identity among minority speakers in Malaysia. Kajian Malaysia, 26(1–2), 1–20.

15. Tashakkori, A., & Teddlie, C. (2010). Mixed methodology: Combining qualitative and quantitative approaches. SAGE Publications.

16. Sandai, R., & Mahmud, C. T. (2023). The influence of cultural attitudes on language development amongst indigenous young Semai children. Southeast Asia Early Childhood Journal, 12(1), 1–10. Retrieved from https://ejournal.upsi.edu.my/index.php/SAECJ/article/view/957

17. Weinreich, U. (1970). Languages in contact: Findings and problems. The Hague: Mouton.

18. Zailani, S. A. B., Makhtar, R., & Yusop, M. S. (2022). Pemilihan bahasa dalam masyarakat Semai di Hulu Selangor. e-Bangi: Journal of Social Sciences and Humanities, 16(9), 115–132. Retrieved from https://ejournals.ukm.my/ebangi/article/view/36294