

Intergenerational Dialect Transmission in Sasak: Evidence from a Geo-Tagged Corpus

Titik Ceriyani Miswaty*, Fahry, Wahyu Kamil Syarifaturrahman

Universitas Bumigora, Mataram, Indonesia

Received: 30th April 2026 | Revised: 15th June 2026 | Accepted: 30th June 2026

*Corresponding author. E-mail: titikceriyani@universitasbumigora.ac.id.

Abstract

The Sasak language of Lombok Island, Indonesia, faces increasing pressure due to urbanization, infrastructure expansion, and growing exposure to national media, raising concerns about intergenerational dialect attrition. Despite growing scholarly attention to Sasak linguistics, no prior study has systematically examined how dialectal features are transmitted across generational cohorts using spatially anchored corpus evidence. This study aims to investigate which phonological, lexical, and morphosyntactic features of Sasak exhibit the highest degree of intergenerational stability, and how geographic location modulates the fidelity of dialect transmission across three generational cohorts. A variationist sociolinguistic approach combined with corpus-based analysis and Geographic Information System (GIS) mapping was employed. Data were collected from 180 native Sasak speakers across 15 geo-tagged villages spanning all five dialect zones, using semi-structured interviews, word-list elicitation, and a dialect perception survey. Acoustic analysis, lexical frequency profiling, and morphosyntactic tagging were conducted using ELAN and Praat. Findings reveal that phonological features consistently exhibit higher intergenerational stability (mean Intergenerational Transmission Index/ITI = 0.78 for Generation 1 to Generation 2; 0.71 for Generation 2 to Generation 3) than lexical features (mean retention = 58.3% for Generation 3). Urban-peripheral zones exhibit accelerated lexical attrition and phonological convergence toward the Ngeno-Ngene prestige dialect, whereas rural-interior communities maintain higher dialectal fidelity. These results underscore the critical role of geographic and social factors in shaping dialect transmission, with implications for language documentation and revitalization in Nusa Tenggara Barat.

Keywords: Dialect Variation; Geo-Tagged Corpus; Intergenerational Transmission; Language Shift; Sasak Language.

How to Cite:

Miswaty, T. C., Fahry, F., & Syarifaturrahman, W. K. (2026). Intergenerational Dialect Transmission in Sasak: Evidence from a Geo-Tagged Corpus. *Humanitatis : Journal of Language and Literature*, 12(2), 243-250.

Copyright ©2026 The Authors.

This article is published by Universitas Bumigora under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

1. INTRODUCTION

Sasak language is the primary vernacular of the indigenous population of Lombok Island, West Nusa Tenggara (NTB), Indonesia. With an estimated 2.9 to 3.2 million speakers, Sasak occupies a prominent position in the linguistic ecology of eastern Indonesia, functioning as both a marker of ethnic identity and a medium of daily communication across diverse social contexts. The language is traditionally classified into five major dialect zones: Kuto-Kute in the north, Ngeto-Ngete in the northeast, Meno-Mene in the central highlands, Ngeno-Ngene

in the south and southwest, and Meriak-Meriku on the southeastern coast, each named after the local variants of the common copula verb *'to be'* (Austin, 2014; Bell, 1984; Coupland, 2007). These dialects differ not only in phonological inventories and prosodic patterns but also in lexical choices and certain morphosyntactic constructions, making Sasak a compelling subject for dialectological inquiry.

Research on Sasak linguistics has grown substantially over the past two decades. Austin (2000b) conducted descriptive fieldwork on Sasak dialectology, documenting phonological distinctions across dialect zones through elicitation methods and finding systematic consonant alternations that serve as reliable dialect boundary markers. Mahyuni (2006) examined speech styles and cultural consciousness in the Sasak community using an ethnographic approach, revealing that register variation is closely tied to social hierarchies and ceremonial contexts. Sugianto and Hasby (2023) investigated dialect leveling in the Mataram urban corridor using a variationist account of Sasak contact, employing structured sociolinguistic interviews with 60 participants, and found that urban-dwelling Sasak speakers progressively shift toward the Ngeno-Ngene prestige variety. Cohn and Ravindranath (2014) studied regional languages in Indonesia, analyzing pressures for language shift through survey and documentary methods across multiple ethnic groups and concluding that bilingual education policies significantly accelerate language attrition among younger cohorts. Campbell and Belew (2020) catalogued endangered languages worldwide using corpus-based documentation methods, finding that geographic isolation strongly predicts intergenerational language maintenance in minority languages. Miswaty et al. (2025) investigated language variation bias in AI-based *text-to-speech* systems through a sociolinguistic study combining corpus analysis and perception experiments, finding that speakers of non-prestige dialects face systematic marginalization in NLP applications. Despite this body of work, the diachronic dimension of dialect maintenance and shift across generational cohorts has received comparatively limited systematic attention, particularly through quantitative, corpus-based methods (Csomay & Crawford, 2024; Farisiyah & Zamzani, 2018; Szmrecsanyi & Rosseel, 2020).

A significant gap exists in understanding how Sasak dialectal features are transmitted or attenuated from older generations to younger speakers, especially as Lombok undergoes rapid urbanization (Kerswill & Wiese, 2022), infrastructure expansion, and increased exposure to national Indonesian media. Classic models of intergenerational language transmission (Eckert, 2012; Errington, 1998; Ewing, 2004) predict that dialect features in contact-dense or high-mobility communities are vulnerable to leveling, whereas geographically isolated communities tend to preserve conservative features. Yet empirical verification of these processes in Sasak (Miswaty et al., 2024, 2025), particularly through geo-spatially anchored corpus evidence, remains absent from the literature.

This study addresses this gap by introducing a geo-tagged spoken corpus of Sasak across three generational cohorts in fifteen village sites spanning all five dialect zones. The overarching aim is to determine (1) which phonological, lexical, and morphosyntactic features exhibit the highest degree of intergenerational stability; (2) how geographic location and proximity to urban centers modulate transmission fidelity; and (3) whether the spatial distribution of dialect features reveals identifiable transmission corridors or rupture zones. By integrating corpus linguistics with geospatial visualization, this study makes both methodological and substantive contributions to the documentation and analysis of Sasak dialectal diversity.

The novelty of this study lies in the first-ever integration of a geo-tagged spoken corpus with Geographic Information System (GIS)-based spatial analysis to systematically quantify intergenerational dialect transmission across three generational cohorts in all five Sasak dialect zones, a methodological approach not previously applied to any regional language in Nusa Tenggara Barat. The contribution of this research is threefold: (1) it produces the first spatially explicit, corpus-based documentation of intergenerational dialect transmission in Sasak; (2) it introduces the Intergenerational Transmission Index (ITI) as a quantifiable and replicable metric for dialect maintenance research; and (3) it identifies geographic transmission corridors and rupture zones with direct implications for targeted language revitalization policy in NTB.

2. RESEARCH METHOD

This study adopts a variationist sociolinguistic framework combined with corpus-based linguistic analysis and Geographic Information System (GIS) mapping. These methods treat linguistic variation as socially structured and statistically measurable rather than random. They are combined with corpus-based analysis, which enables objective, large-scale quantification of feature frequencies across a large number of speakers, and integrated with Geographic Information System (GIS) mapping, which geo-tags each data point with GPS coordinates so that patterns of dialect retention and loss can be visualized spatially across the island. This integration allows the research to simultaneously determine the extent to which each dialect feature changes across generations, which social factors predict that change, and where geographically the change is concentrated. The variationist approach is particularly appropriate for this study because it enables systematic quantification of linguistic variables across social and geographic dimensions simultaneously (Labov, 1973, 1997), rather than relying on impressionistic description. This framework has been widely validated in dialect transmission research (Diegoli, 2024; He & Wang, 2013; Van Dorst et al., 2024), making it methodologically robust for investigating intergenerational change. Corpus-based analysis was chosen because it allows objective, replicable measurement of feature frequency and distribution across large datasets, thereby minimizing observer-effect bias. GIS integration was incorporated because of its value for geospatial visualization in documenting spatial patterns of linguistic change in low-resource Indo-Pacific languages. The research design is cross-generational and simultaneous (*apparent-time design*), capturing speakers from three age cohorts across geographically distributed field sites (Budiono & Yanita, 2024; He & Wang, 2013; Oliver, 2022).

The study site encompasses fifteen villages distributed across the five traditional Sasak dialect zones on Lombok Island. Three villages were selected per dialect zone based on the criteria of dialectal representativeness (confirmed by preliminary surveys), generational community composition (the presence of all three target age cohorts), and geographic diversity (with coastal, peri-urban, and inland sites represented within each zone).

Participants were 180 native Sasak speakers divided into three generational cohorts: Generation 1 (G1, ages 60–75, $n = 60$), Generation 2 (G2, ages 35–50, $n = 60$), and Generation 3 (G3, ages 17–30, $n = 60$). Participants were recruited through purposive sampling with the assistance of local village heads. All participants were lifelong residents of their respective villages, monolingual speakers or dominant-Sasak bilinguals, and had not resided outside their dialect zone for more than six cumulative months. Informed consent was obtained from all participants in line with research ethics guidelines.

Data collection employed three instruments: (1) a 45-minute semi-structured conversational interview conducted entirely in Sasak, covering topics such as agriculture, family, village life, and local traditions; (2) a word-list elicitation task comprising 250 lexical items drawn from the Swadesh list and supplemented by domain-specific Sasak vocabulary; and (3) a dialect perception survey using recorded stimuli representing all five dialect zones. Each session was audio-recorded using a Zoom H5 field recorder and geo-tagged with GPS coordinates using a Garmin eTrex 32x. These instruments were triangulated to ensure convergent validity across the production, recognition, and metalinguistic dimensions of dialect competence.

The recorded audio was transcribed orthographically and phonetically by trained transcribers who were native Sasak speakers with linguistic training, using ELAN v6.4 annotation software. Each utterance tier was linked to GPS metadata. Transcriptions were verified for accuracy through a double-blind inter-rater reliability check, with a minimum Cohen's $\kappa = 0.82$ required before inclusion in the final corpus.

Linguistic analysis proceeded on three levels. First, phonological variables (e.g., the realization of final /q/ in Kuto-Kute, vowel height in Meriak-Meriku interrogatives, and the /e/-/ɛ/ contrast in Meno-Mene) were analyzed using Praat v6.3 for acoustic measurements and logistic regression models in R (v4.3) for statistical analysis. Second, lexical retention was quantified through frequency profiling in AntConc v4.2 using dialect-specific reference word lists. Third, morphosyntactic features were tagged using custom Python scripts (the spaCy framework adapted for Sasak) and analyzed using mixed-effects models controlling for participant age, education,

and geographic zone. An Intergenerational Transmission Index (ITI) was calculated for each linguistic variable as the correlation coefficient between the feature frequencies of consecutive generational cohorts within the same village.

3. FINDINGS AND DISCUSSION

The analysis of the geo-tagged Sasak corpus yielded rich quantitative and spatial evidence concerning three primary linguistic dimensions: phonological feature transmission, lexical retention, and morphosyntactic stability. Before presenting the findings, several statistical abbreviations used throughout this section are defined as follows: SD refers to Standard Deviation, a measure of the dispersion of data values around the mean; G1, G2, and G3 refer to Generation 1 (ages 60–75), Generation 2 (ages 35–50), and Generation 3 (ages 17–30), respectively; *t* refers to the *t*-statistic from an independent-samples *t*-test used to compare group means; *p* refers to the *p*-value indicating statistical significance (values below .05 are considered statistically significant); and β refers to the standardized regression coefficient in logistic regression models. Each linguistic dimension is treated in a dedicated subsection below, followed by an integrative discussion of the observed spatial patterns.

3.1. Phonological Feature Transmission Across Generations

Phonological features demonstrated the strongest intergenerational stability among the three dimensions analyzed. Across all five dialect zones, the mean Intergenerational Transmission Index (ITI) for the seven targeted phonological variables averaged 0.78 (SD = 0.09) for G1-to-G2 transmission and 0.71 (SD = 0.12) for G2-to-G3 transmission, on a scale ranging from 0 (complete loss) to 1.0 (complete retention). These values indicate that the majority of phonological distinctiveness is preserved across generations, although a statistically significant decline was observed between G2 and G3 in peri-urban sites ($t(58) = 3.22, p = .002$).

The most stable phonological variable was the realization of the final glottal stop /q/ in the Kuto-Kute dialect, which maintained an ITI = 0.91 across all three cohorts, even in peri-urban areas of Northern Lombok (Sugianto & Hasby, 2023; Trudgill, 1983). This robustness may reflect the social salience of this feature as a marker of local identity. Conversely, the /e/-/ɛ/ vowel height contrast in Meno-Mene showed significant erosion among G3 speakers (ITI = 0.54), particularly among those with regular exposure to Indonesian national television and higher levels of secondary education, factors consistent with *koineization* pressures documented in other minority-language contexts (Kerswill & Wiese, 2022; Trudgill, 1986).

Acoustic analysis confirmed a gradual compression of the vowel space among G3 urban-peripheral speakers relative to G1 rural speakers, particularly with respect to the front-mid vowel contrast. GIS mapping revealed that the 0.70 ITI contour line for the /e/-/ɛ/ contrast closely follows the 10km urban boundary surrounding Mataram, suggesting that phonological convergence is geographically circumscribed rather than island-wide, a pattern consistent with the gravity model of dialect diffusion (Trudgill, 2020).

Table 1. Intergenerational Transmission Index (ITI) by Dialect Zone and Cohort Transition

Dialect Zone	G1 → G2 ITI	G2 → G3 ITI (Rural)	G2 → G3 ITI (Peri-Urban)
Kuto-Kute	0.88	0.83	0.74
Ngeto-Ngete	0.81	0.74	0.65
Meno-Mene	0.79	0.71	0.58
Ngeno-Ngene	0.72	0.67	0.61
Meriak-Meriku	0.76	0.70	0.59
Man	0.79	0.73	0.63

Table 1 above presents the Intergenerational Transmission Index (ITI) values disaggregated by dialect zone and cohort transition, further subdivided into rural and peri-urban subcategories for the G2→G3 interval. The data reveal a consistent pattern across all five zones: ITI values decline monotonically from the G1→G2 transition to the G2→G3 rural transition and decline further in peri-urban G2→G3 settings. The most stable zone is

Kuto-Kute ($G1 \rightarrow G2$ ITI = 0.88; $G2 \rightarrow G3$ rural ITI = 0.83), reflecting the geographic isolation of northern Lombok communities, whereas the lowest ITI values are recorded in the peri-urban sites of Meno-Mene and Meriak-Meriku (ITI = 0.58 and 0.59, respectively), indicating that phonological erosion is most pronounced in communities with high levels of exposure to the urban Mataram corridor. The mean ITI across all zones declines from 0.79 ($G1 \rightarrow G2$) to 0.73 ($G2 \rightarrow G3$ rural) and 0.63 ($G2 \rightarrow G3$ peri-urban), confirming that urbanization exerts a measurable and spatially stratified influence on phonological transmission.

3.2. Lexical Retention and Dialect-Specific Vocabulary Loss

Lexical features exhibited considerably greater intergenerational attrition than phonological variables. Analysis of the 250-item word list revealed that dialect-specific lexical items, defined as those with documented variants differing across at least three dialect zones, showed a mean G3 retention rate of 58.3% (SD = 14.7%), compared to 81.6% in G1 (SD = 7.2%). The steepest losses were observed in the cultural and agricultural vocabulary domains, where Sasak-specific terms for traditional irrigation, rice cultivation stages, and kinship terminology showed G3 retention rates below 40% in peri-urban sites (see Table 1). As shown in Table 1, every dialect zone exhibits a declining trajectory from G1–G2 to G2–G3 transmission, with peri-urban zones consistently scoring lower than their rural-interior counterparts.

The lexical frequency profile analysis indicated that G3 speakers increasingly substitute dialect-specific Sasak lexemes with Indonesian equivalents or adopt regional Lombok Malay forms, particularly in domains mediated by formal education and social media. This finding aligns with Boyd's (2010) observation on heritage-language lexical attrition in diaspora contexts and extends it to intra-island dialect contact. Interestingly, the Ngeno-Ngene dialect, which most closely approximates the prestige standard, showed the lowest lexical attrition rate among G3 speakers (42.1%) (Thieberger, 2022), suggesting that speakers of prestige dialects experience less external lexical pressure than speakers of lower-prestige peripheral dialects such as Kuto-Kute (G3 retention rate = 32.6%).

The geo-tagged corpus enabled the spatial tracking of *lexical isoglosses*, geographic boundaries separating communities that retain versus lose specific dialect lexemes. Kernel density analysis in QGIS revealed that lexical isoglosses among G3 speakers have shifted inward toward geographically peripheral zones compared to those of G1 speakers, indicating that the territorial distribution of dialect-unique vocabulary has contracted. This pattern suggests that lexical retreat is not uniform but is concentrated along transportation arterials, consistent with the route-based diffusion model (Labov, 2001).

3.3. Morphosyntactic Stability and Generational Variation

Morphosyntactic features exhibited intermediate stability, situated between the high retention of phonological features and the marked attrition of lexical features. The verbal affixation system, particularly the N- active voice prefix and the -an applicative suffix, maintained relatively high retention across all cohorts (G1: 94.2%, G2: 89.7%, G3: 84.3%), suggesting that core grammatical mechanisms are more resistant to intergenerational change than surface lexical forms. These findings resonate with Myers-Scotton's (2002) prediction that morphological frames are the last elements to be replaced in language-contact situations. More variation was observed in non-core morphosyntactic features, particularly copula realization. The five-dialect copula system (kuto/ngeto/meno/ngeno/meriak) showed significant simplification among G3 speakers, with 34.7% of G3 speakers across all sites defaulting to the Ngeno-Ngene copula form even when speaking in their home dialect zone. This convergence toward the Ngeno-Ngene standard was statistically predicted by (a) frequency of visits to Mataram ($\beta = 0.41$, $p < .001$), (b) years of formal education ($\beta = 0.27$, $p = .003$), and (c) frequency of Indonesian social media use ($\beta = 0.33$, $p < .001$) in a multivariate logistic regression model (Nagelkerke $R^2 = 0.52$). Future researchers are encouraged to extend the geo-tagged corpus methodology to include written vernacular texts and social media language data in Sasak, which may reveal additional transmission pathways beyond naturalistic speech. Longitudinal re-sampling of the same participants over a period of five to ten years would strengthen the apparent-time findings with real-time evidence of linguistic change. Additionally, comparative studies across other regional languages of Nusa Tenggara Barat

(such as Bima, Sumbawa, and Manggarai) would help determine whether the patterns identified here are specific to Sasak or reflect broader sociolinguistic dynamics in eastern Indonesia. Interventions for dialect documentation and revitalization should prioritize lexical domains and peripheral geographic zones, where attrition is most advanced.

These results collectively indicate that intergenerational dialect transmission in Sasak is not a uniform process but is modulated by the interaction of geographic location, social mobility, and exposure to prestige varieties. Rural-interior communities in the Kuto-Kute and Meno-Mene zones maintain the highest transmission fidelity across all three linguistic dimensions, functioning as reservoir zones for dialect preservation. Peri-urban zones, particularly those along the Mataram–Selong transportation corridor, exhibit accelerated convergence, constituting what Trudgill (2004) terms ‘*new dialect*’ formation zones. The geo-tagged corpus provides the first spatially explicit evidence of this pattern in Sasak, offering a replicable methodological model for endangered-language documentation in archipelagic Southeast Asia.

In summary, the findings of this study both corroborate and extend prior research on minority-language transmission. The pattern of phonological stability paired with accelerated lexical attrition is consistent with Myers-Scotton’s (2002) prediction that grammatical and phonological frames are the last systems to be replaced in language-contact situations and aligns with Sugianto and Hasby (2023) documentation of Ngeno-Ngene prestige diffusion in urban Lombok. However, the present study departs from earlier generalist accounts by demonstrating that dialect loss in Sasak is not island-wide but spatially circumscribed within a 10km radius of Mataram, a finding that challenges the assumption of uniform language shift and instead reveals a mosaic of transmission zones with distinct trajectories. This study therefore both supports previous research on the social and geographic determinants of dialect maintenance and significantly refines that understanding by introducing geospatial evidence that was previously absent from the Sasak sociolinguistic literature. These results carry practical implications: revitalization and documentation efforts should be spatially targeted, prioritizing peri-urban lexical domains and the Meno-Mene and Meriak-Meriku zones, where attrition is most advanced and the risk of irreversible dialect loss is greatest.

4. CONCLUSION

This study has demonstrated that intergenerational dialect transmission in Sasak is a geographically stratified and feature-variable process in which phonological features are transmitted with greater fidelity than lexical or morphosyntactic features, and in which rural-interior communities consistently outperform peri-urban communities in dialect maintenance. Across all five dialect zones, phonological ITI scores averaged 0.78 (G1→G2) and 0.71 (G2→G3 rural), whereas lexical retention declined to 58.3% among G3 speakers. The geo-tagged corpus methodology enabled the spatial tracking of dialect transmission corridors and rupture zones, revealing that urbanization exerts a spatially bounded influence that intensifies within a 10km radius of Mataram. The Ngeno-Ngene variety is expanding its influence beyond its traditional boundaries, accelerating copula-system simplification across all zones, while the Kuto-Kute glottal stop remains the most socially salient and resistant feature.

Future researchers are encouraged to extend the geo-tagged corpus methodology to include written vernacular texts and social media language data in Sasak, which may reveal additional transmission pathways beyond naturalistic speech. Longitudinal re-sampling of the same participants over a period of five to ten years would strengthen the *apparent-time* findings with real-time evidence of linguistic change. Additionally, comparative studies across other regional languages of Nusa Tenggara Barat would help determine whether the patterns identified here are specific to Sasak or reflect broader sociolinguistic dynamics in eastern Indonesia. Interventions for dialect documentation and revitalization should prioritize lexical domains and peripheral geographic zones, where attrition is most advanced.

REFERENCES

- Austin, P. K. (2014). Tense, aspect, mood and evidentiality in Sasak, eastern Indonesia. *Language Documentation and Description*, 231–251. <https://doi.org/10.25894/LDD182>

- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204. <https://doi.org/10.1017/S004740450001037X>
- Budiono, S., & Yanita, S. R. (2024). Language Documentation Practices in Indonesia: A Critical Theory Perspective. *Lingua Cultura*, 18(1), 133–140. <https://doi.org/10.21512/lc.v18i1.11878>
- Campbell, L., & Belew, A. (Eds.). (2020). *Cataloguing the world's endangered languages*. Routledge. OCLC: 1156996572.
- Cohn, A. C., & Ravindranath, M. (2014). Local Languages in Indonesia: Language Maintenance or Language Shift? *Linguistik Indonesia*, 32(2), 131–148. <https://doi.org/10.26499/li.v32i2.22>
- Coupland, N. (2007). *Style*. Cambridge University Press. OCLC: 1481413251.
- Csomay, E., & Crawford, W. J. (2024, January 22). *Doing Corpus Linguistics* (2nd ed.). Routledge. <https://doi.org/10.4324/9781003363309>
- Diegoli, E. (2024). A corpus-assisted analysis of indexical signs for (im)politeness in Japanese apology-like behaviour. *Journal of Politeness Research*, 20(2), 427–453. <https://doi.org/10.1515/pr-2022-0002>
- Eckert, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology*, 41, 87–100. <https://doi.org/10.1146/annurev-anthro-092611-145828>
- Errington, J. J. (1998). *Shifting languages : Interaction and identity in Javanese Indonesia*. Cambridge ; New York : Cambridge University Press. <http://archive.org/details/shiftinglanguage0000erri>
- Ewing, M. C. (2004). Colloquial Indonesian. In K. A. Adelaar & N. Himmelmann (Eds.), *The Austronesian Languages of Asia and Madagascar*. Routledge.
- Farisiyah, U., & Zamzani, Z. (2018). Language Shift and Language Maintenance of Local Languages toward Indonesian. *Proceedings of the International Conference of Communication Science Research (ICCSR 2018)*. <https://doi.org/10.2991/iccsr-18.2018.50>
- He, Y., & Wang, H. (2013). A Corpus-Based Study of Epistemic Modality Markers in Chinese Research Articles. In D. Ji & G. Xiao (Eds.), *Chinese Lexical Semantics* (pp. 199–208). Springer. https://doi.org/10.1007/978-3-642-36337-5_21
- Kerswill, P., & Wiese, H. (Eds.). (2022, March 30). *Urban Contact Dialects and Language Change: Insights from the Global North and South*. Routledge. <https://doi.org/10.4324/9780429487958>
- Labov, W. (1973). *Sociolinguistic patterns*. Philadelphia, University of Pennsylvania Press. <http://archive.org/details/sociolinguisticp00will>
- Labov, W. (1997). The Social Stratification of (r) in New York City Department Stores. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics: A Reader* (pp. 168–178). Macmillan Education UK. https://doi.org/10.1007/978-1-349-25582-5_14
- Labov, W. (2001, March 30). *Principles of Linguistic Change, Social Factors*. Wiley.
- Mahyuni, M. (2006). *Speech Styles and Cultural Consciousness in Sasak Community*. Yayasan Cerdas.
- Miswaty, T. C., Fahry, Triwijoyo, B. K., & Patty, E. N. S. (2025). Mitigating Language Variation Bias in AI-Based Text-to-Speech Systems: A Sociolinguistic Study in Natural Language Processing. *2025 7th International Conference on Cybernetics and Intelligent System (ICORIS)*, 1–5. <https://doi.org/10.1109/ICORIS67789.2025.11296112>
- Miswaty, T. C., Widhi, B. A., Alawiyah, R., Elhia, M. W., & Nuralfilail, N. (2024). Investigating Regional Variation and Digital Dialect in Online Game Identity. *Journal of Languages and Language Teaching*, 12(3), 1366. <https://doi.org/10.33394/jollt.v12i3.11600>

- Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford University Press.
- Oliver, S. J. (2022). A corpus-based approach to (im)politeness metalanguage: A case study on Shakespeare's plays. *Journal of Pragmatics*, 199, 6–20. <https://doi.org/10.1016/j.pragma.2022.07.001>
- Sugianto, S., & Hasby, M. A. (2023). The Language Variations Used among Speakers of Sasak Lombok, West Nusa Tenggara. *JOLLT Journal of Languages and Language Teaching*, 11(1), 144–150. <https://doi.org/10.33394/jollt.v11i1.6071>
- Szmrecsanyi, B., & Rosseel, L. (2020). English Corpus Linguistics. In B. Aarts, A. McMahon, & L. Hinrichs (Eds.), *The Handbook of English Linguistics* (1st ed., pp. 29–44). Wiley. <https://doi.org/10.1002/9781119540618.ch3>
- Trudgill, P. (1983). *On dialect: Social and geographical perspectives*. New York : New York University Press. <http://archive.org/details/ondialectsocialg0000trud>
- Trudgill, P. (1986). *Dialects in contact*. B. Blackwell.
- Trudgill, P. (2004). *Dialects*. Psychology Press.
- Trudgill, P. (2020). *Millennia of language change: Sociolinguistic studies in deep historical linguistics*. Cambridge university press.
- Van Dorst, I., Gillings, M., & Culpeper, J. (2024). Sociopragmatic variation in Britain: A corpus-based study of politeness. *Journal of Pragmatics*, 227, 37–56. <https://doi.org/10.1016/j.pragma.2024.04.009>