Prosodi: Jurnal Ilmu Bahasa dan Sastra
Vol. 20 – No. 2, Oktober 2026, pp.305 - 319
https://doi.org/10.21107/prosodi.v20i2.36454
Open Access:
https://journal.trunojoyo.ac.id/prosodi

ISSN 1907-6665 (Print)
ISSN 2622-0474 (Online)

LEXICAL VARIATION OF MADURESE IN BANGKALAN: GEOGRAPHIC PATTERNS AND POSSIBLE AGE-RELATED DIFFERENCESS

Khairatul Amaliya[1], Zulkhaeriyah Zulkhaeriyah[2*], Diah Ekawati Ayuningtias[3], Desi Puspitasari[4] & Agustinus GR Dasion[5]
[1,2,3,4,5]Universitas Trunodjoyo Madura, Indonesia
**Corresponding Author:* zulkhaeriyah@trunojoyo.ac.id

Article History

Received : 08/10/2026
Revised : 14/10/2026
Accepted : 30/10/2026
Published : 02/11/2026

Keywords:
Age, Bangkalan, Geography, Lexical Variation.

Kata Kunci:
Umur, Bangkalan, Geografi, Variasi Bahasa.

Abstract. This study examines how lexical variation in Bangkalan Madurese is distributed across geography and age cohorts. Data were collected in five *kecamatan* (sub-districts) of Bangkalan Regency through a 40-item word-list questionnaire completed by 15 informants from three age cohorts and through semi-structured interviews with seven informants, whose transcripts were analyzed thematically. Geographic variation was the clearest pattern in the sample: the second-person pronoun *kamu* "you", for example, is *kakeh* in Tanjung Bumi, *hedeh* in Labang, and *be'na* in Kota Bangkalan and Tanah Merah in the age cohorts represented in each *kecamatan*. Possible age-related contrasts were few and small. The most notable tentative signal concerned the high speech register (*èngghi-bhunten*), which only one of fifteen informants produced in the questionnaire and which interviewees described as not fully mastered. The interviews also contained individual reports of lexical loss (*obu'* "hair"), of colour terms replaced by loanwords, and of differences in self-reported Madurese use between communities; these reports are exploratory. Because the sample is small and unevenly distributed across *kecamatan* and age cohorts, the findings are preliminary. They suggest that geography shapes everyday vocabulary, whereas age-related differences, register competence, and lexical change require further investigation with larger and more balanced samples.

Abstrak. Penelitian ini mengkaji bagaimana variasi leksikal bahasa Madura di Bangkalan tersebar menurut geografi dan kelompok usia. Data dikumpulkan di lima kecamatan di Kabupaten Bangkalan melalui kuesioner daftar kata 40 butir yang diisi oleh 15 informan dari tiga kelompok usia dan melalui wawancara semi-terstruktur dengan tujuh informan, yang transkripnya dianalisis secara tematik. Variasi geografis merupakan pola yang paling jelas dalam sampel: pronomina persona kedua *kamu*, misalnya, adalah *kakeh* di Tanjung Bumi, *hedeh* di Labang, dan *be'na* di Kota Bangkalan dan Tanah Merah pada kelompok usia yang terwakili di setiap kecamatan. Perbedaan yang mungkin terkait usia sedikit dan kecil. Sinyal tentatif yang paling menonjol terdapat pada tingkat tutur tinggi (*èngghi-bhunten*), yang hanya dihasilkan oleh satu dari lima belas informan dalam kuesioner dan yang oleh informan wawancara dinyatakan tidak dikuasai sepenuhnya. Wawancara juga memuat laporan individual tentang hilangnya kosakata (*obu'* "rambut"), penggantian istilah warna dengan kata serapan, serta perbedaan penggunaan bahasa Madura yang dilaporkan sendiri antarkomunitas; laporan ini bersifat eksploratif. Karena sampel kecil dan tidak tersebar merata menurut kecamatan dan kelompok usia, temuan ini bersifat awal. Temuan ini menunjukkan bahwa geografi membentuk kosakata sehari-hari, sedangkan perbedaan terkait usia, penguasaan tingkat tutur, dan perubahan leksikal memerlukan kajian lebih lanjut dengan sampel yang lebih besar dan lebih seimbang.

How to Cite: Amaliya, K., Zulkhaeriyah, Z., Ayuningtias, D.E., Puspitasari, D., Dasion, A.GR. (2026). LEXICALVARIATION OF MADURESE IN BANGKALAN: GEOGRAPHIC PATTERNS AND AGE-RELATED CHANGE. *Prosodi: Jurnal Ilmu Bahasa dan Sastra*, 20(2), 305-319.
https://doi.org/10.21107/prosodi.v20i2.36454

Correspondence address:

Correspondence ORCID ID: 0000-0002-6512-8115

Publisher:
Program Studi Sastra Inggris, Universitas Trunodjoyo Madura. Jln. Raya Telang, PO Box 2 Telang, Kamal, Bangkalan, Madura.
prosodi@trunojoyo.ac.id

INTRODUCTION

Indonesia is a multicultural and multilingual nation, home to a wide array of ethnic groups, traditions, and more than 700 local languages spread across its islands (Eberhard et al., 2023). This linguistic diversity is not static. Every regional language exhibits internal variation shaped by geography, social class, education, gender, and, as this study focuses on, generational age (Labov, 1972).

Previous studies have shown that language variation can be examined through both geographic and social dimensions. In the context of Madurese, Muttaqin, Sahiruddin, and Rodliyah (2019) investigated lexical and syntactic variation across regions and age groups in Madura, using a written instrument completed by 90 participants from three regencies. Their findings showed that lexical differences occurred among speakers and were related to the speakers' socio-cultural background, while politeness levels also influenced lexical choices across age groups and regions. In a more geographically focused study, Sugiharti and Pariyanto (2021) examined lexical variation in Madurese spoken in Bangkalan Regency using 335 glosses distributed across twelve semantic fields. Their study identified substantial lexical differences among the research areas in Bangkalan, showing that lexical variation exists even within the same regency. Taken together, these studies indicate that Madurese variation cannot be understood only as a difference between broad dialect areas, but also involves social factors such as age and politeness level.

Research on other regional languages in Indonesia points to similar tensions between regional and social variation, and it allows a comparison of how speech levels are maintained in different communities. For Javanese, Puspandari (2026) compared lexical, phonological, morphological, and sociolinguistic features across three regions of East Java and showed that Javanese within one province cannot be treated as uniform. Studies of Javanese speech levels, in turn, describe a shift away from formal styles toward Indonesian among younger speakers, linked to new educational and social opportunities (Errington, 1998);(Smith-Hefner, 2009). A comparable concern has been raised for Balinese: Sutika (2019) reports that the speech levels (*anggah-ungguhing basa*) are still used on traditional and formal occasions and when speaking with elders and strangers, but that their use has decreased because young people are less interested in them and prefer Indonesian. Comparing Madurese with these languages is useful for two reasons. It shows that a weakening of polite registers, if present in Madurese, would not be unique to one language; and it shows that evidence for such weakening is of a different kind from evidence for regional lexical variation, and so should be weighed separately.

Madurese is one of Indonesia's major regional languages and is widely spoken on Madura Island and in many regions of East Java. It possesses a complex speech-level system that reflects the social structure of Madurese society: *enjâ'-iyâ* (low/informal level), *èngghi-ènten* (middle, polite level), and *èngghi-bhunten* (high, very polite level) (Davies, 2010). Bangkalan Regency lies at the westernmost tip of Madura Island and, like any living language, Bangkalan Madurese is not homogeneous. During fieldwork in five *kecamatan* of Bangkalan, informants repeatedly volunteered that vocabulary and pronunciation differ not only between *kecamatan* but even between neighboring villages. Several informants also reported that the high speech register (*èngghi-bhunten*) is increasingly unfamiliar to younger speakers, and in some cases to middle-aged speakers as well. These are informants' own observations; the present study treats them as questions to be examined against the word-list data, not as established facts.

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Speech-level variation is another important dimension of Madurese linguistic practice. Madurese is commonly described as having three speech levels: Basa Enjâ'-iyâ, Basa Èngghi-ènten, and Basa Èngghi-Bhunten. These levels are associated with differences in social relationships, status, and politeness, and speakers may select different lexical forms depending on the interlocutor and communicative situation (Afiansyah, 2023). Therefore, lexical variation in Madurese does not always indicate a geographic difference. A speaker may use different lexical forms because of differences in social context or the level of politeness required in an interaction. This distinction is particularly relevant to the present study because some lexical items in the word list are associated with higher speech levels. Examining these forms together with everyday vocabulary makes it possible to consider whether differences found among speakers are primarily geographic, age-related, or connected to speech-level use.

Local languages now face increasing pressure from globalization and modernization. The dominance of Indonesian as the national language (Sneddon, 2003), combined with social stigma toward Madurese in wider inter-ethnic settings (several informants used the word *minder*, "feeling inferior/self-conscious"), may affect community attitudes toward, and use of, the language. Such pressures are the kind of conditions under which intergenerational transmission can weaken and language shift can begin (Fishman, 1991).

Previous studies have established that Madurese shows lexical variation across regions and social groups, but they leave several questions open. First, Muttaqin et al. (2019) worked at the level of regencies and used a written instrument, so their data cannot show how regional and age-related patterns interact within the smaller units of a single regency. Second, Sugiharti and Pariyanto (2021) documented lexical variation within Bangkalan across a large set of glosses, but their focus, as described above, was the distribution of forms across research areas rather than a comparison of age cohorts within the same sub-district. Third, work on Javanese and Balinese speech levels (Errington, 1998);(Smith-Hefner, 2009); (Sutika, 2019)discusses the weakening of polite registers among younger speakers, but to our knowledge no comparable study of Bangkalan Madurese examines register competence together with lexical variation in the same group of speakers. The present study addresses this gap in a limited way. It examines lexical forms elicited with a 40-item word list in five sub-districts of Bangkalan, compares them across three age cohorts, and uses interviews to explore how speakers describe the differences they notice. Rather than assuming that younger speakers necessarily show more change, the study asks whether differences between age cohorts remain visible once geographic variation is taken into account. Its contribution is therefore exploratory: a first sub-district-level comparison of geographic and possible age-related patterns in Bangkalan Madurese, based on a small sample, which indicates where larger studies should look. The question also matters for practical reasons. If lexical differences are primarily geographic, documentation and teaching must take local varieties into account rather than assume a single standard. If some differences prove to be generational, they would indicate where transmission may be weakest and where maintenance efforts may be most urgent.

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Three concepts frame the analysis. First, dialectology describes and maps variation across geographic and social space, and the word-list questionnaire is its long-established instrument: presenting every informant with the same concepts yields directly comparable forms, and sampling across age groups allows the same data to speak to both geographic and social variation (Tagliamonte, 2012). Second, the apparent-time construct treats differences among age cohorts at a single point in time as indexing change across generations (Tagliamonte, 2012), which is best interpreted together with the social structure of the speech community (Weinreich et al., 1968). Age differences may, however, reflect age grading rather than generational change (Holmes & Wilson, 2017), and the two can be separated strictly only with real-time evidence, so apparent-time findings from a limited sample are informative but provisional. Because command of polite registers is acquired through socialization and use, as in the speech levels of Javanese (Holmes & Wilson, 2017), register competence is a domain where generational and sociocultural factors interact closely. Third, lexical borrowing is among the most visible indicators of language contact and often the first sign of broader shift (Thomason & Kaufman, 1988). Borrowing need not signal decline, but it becomes a concern for vitality when loanwords systematically replace core inherited vocabulary and younger speakers no longer acquire the inherited terms (Fishman, 1991).

In the present study, these concepts are used as complementary perspectives rather than as separate explanations of language variation. Geographic variation is identified by comparing lexical forms across the five selected sub-districts, while age-related variation is considered when speakers from different age groups within the same location use different forms. The apparent-time perspective is used cautiously because differences between age groups may reflect either generational change or age grading rather than permanent language change. Speech-level variation is examined through the use and recognition of forms associated with higher levels of Madurese, particularly *èngghi-bhunten*. Finally, lexical borrowing and replacement are considered when speakers report or demonstrate the use of Indonesian or other contemporary forms in place of older Madurese vocabulary. By combining these perspectives, the study does not treat every difference between speakers as evidence of language change. Instead, it first considers the distribution of each form across locations and age groups before interpreting the possible social meaning of the variation.

Madurese is generally divided into dialect areas roughly corresponding to the four regencies of Madura Island (Davies, 2010). Within Bangkalan, informants repeatedly observed that *setiap daerah berbeda* “every area is different”, and this study examines whether the word-list data support that observation. The combination of a structured word list and semi-structured interviews is useful for this purpose. The word list provides comparable lexical data across speakers and locations, making it possible to identify recurring geographic patterns and possible differences between age groups. The interviews provide contextual information that isolated lexical items cannot: speakers can explain where particular forms are used, whether certain words are still familiar to younger speakers, and how they perceive the use of different speech levels. The word list thus shows patterns of lexical distribution, while the interviews offer speakers' own accounts of those patterns. Because the interview sample is small, these accounts are treated as exploratory and illustrative, not as evidence of community-wide patterns. The study addresses two research questions:

(1) To what extent does lexical variation in the word-list data pattern along geographic (*kecamatan*) lines rather than along age cohorts?

(2) Which possible age-related differences appear in the word-list and interview data (in specific vocabulary items and in use of the high speech register), and how far can they be interpreted as indicating change?

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

RESEARCH METHODS

This study employs a descriptive qualitative method that combines a structured lexical questionnaire with semi-structured interviews, in line with standard dialect-survey methodology (Chambers & Trudgill, 1998). The word list yields comparable lexical forms for a fixed set of concepts, whereas the interviews yield informants' own accounts of how their language use has changed and how they perceive differences between areas and generations. Using both makes it possible to compare what speakers report with what they actually produced.

The qualitative approach was selected because the study aims not only to identify differences in lexical forms but also to understand how speakers use and evaluate those forms in their everyday communication. The structured lexical questionnaire provides comparable data across speakers and locations, while the semi-structured interviews provide additional information about language use, perceptions, and experiences that cannot be captured through the word list alone. This combination allows the study to examine lexical variation from both a distributional and a contextual perspective. The questionnaire helps identify recurring forms across the five sub-districts and three age groups, whereas the interviews help explain possible reasons behind the continued use, reduced use, or replacement of particular lexical items. Therefore, the two instruments were treated as complementary sources of evidence rather than as separate datasets.

Research Sites and Participants

Data were collected in five *kecamatan* of Bangkalan Regency: Kota Bangkalan, Labang, Tanah Merah, Kokop, and Tanjung Bumi. The five sub-districts are those covered by a lecturer's research project within which the data were collected; they therefore capture lexical variation across different parts of Bangkalan Regency, but they were not chosen to represent the regency as a single homogeneous speech community. The locations provide a basis for comparing lexical forms across speakers who live in different geographic areas while using the same structured word list. Comparing the five locations is relevant because differences in Madurese vocabulary may occur not only between major dialect regions but also at a more local level. Informants were recruited through purposive sampling, on the criteria that they were native speakers of Madurese and had lived in the *kecamatan* concerned for a considerable length of time. A 40-item word-list questionnaire was completed by 15 informants, and 7 semi-structured interviews were transcribed and analyzed (further interviews remain to be transcribed and are not included here). Six of the seven interviewees also completed the word list, allowing direct triangulation between their questionnaire responses and their own reflections on language use. Informants were grouped into three age cohorts: older (55+), middle (35–54), and younger (17–34). To protect their identity, informants are not named but identified by a code made up of the initials of their name and their age (e.g., AM-61 is a 61-year-old informant). The same code is used for an informant in both instruments.

The distribution of informants is uneven (Table 1). Each *kecamatan* is represented by only two to four word-list informants, and no *kecamatan* has informants from all three age cohorts, so every comparison between cohorts within a *kecamatan* involves only two of the three. Because location and age are partly confounded in this way, the geographic and age-related comparisons are not fully independent, and the results are reported as patterns in this sample, not as representative of each *kecamatan* or of Bangkalan as a whole. The consequences for interpretation are taken up in the Findings and Discussion and in the Limitations.

Table 1. Word-list informants by kecamatan and age cohort

Kecamatan	Older (55+)	Middle (35–54)	Younger (17–34)
Kota Bangkalan	AM-61, MZ-58	AS-38	–
Labang	SU-56	–	AR-30, HE-25

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Kecamatan	Older (55+)	Middle (35–54)	Younger (17–34)
Tanah Merah	–	MO-35	DH-33
Kokop	–	SI-37, SR-52	SE-20
Tanjung Bumi	–	AH-50, MA-35	MU-31, SA-33

Note: the older cohort is not represented in Tanah Merah, Kokop, or Tanjung Bumi; the middle cohort is not represented in Labang; and the younger cohort is not represented in Kota Bangkalan (see Limitations).

Table 2. Interview informants (transcribed and analyzed)

Code	Age	*Kecamatan*	Cohort	Also word list?
AB-30	~30	Tanah Merah	Younger	No
MU-31	31	Tanjung Bumi	Younger	Yes
DH-33	33	Tanah Merah	Younger	Yes
SI-37	37	Kokop	Middle	Yes
SE-20	20	Kokop	Younger	Yes
HE-25	25	Labang	Younger	Yes
AM-61	61	Kota Bangkalan	Older	Yes

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Instruments

- A structured word-list questionnaire covering five domains (body parts, pronouns/demonstratives, nature/environment, basic verbs, and adjectives/numerals; 40 items in total) was completed by all 15 informants. The word list was administered in person by the researcher, who was present during each session. The concepts were presented in Indonesian and informants gave their responses in Madurese. The sessions were audio-recorded with a mobile phone and the researcher transcribed the responses. The items were chosen to represent these basic vocabulary domains so that comparable lexical forms could be elicited from informants across the five sub-districts and three age groups.

The word-list questionnaire was designed to provide a consistent set of lexical prompts for all informants. Using the same 40 concepts across the five locations allowed the researchers to compare the forms produced by speakers without changing the elicitation items from one location to another. The selected domains cover basic and frequently encountered vocabulary, including terms for body parts, personal reference, the natural environment, everyday actions, and descriptive or numerical concepts. These categories were used to obtain a broad picture of everyday lexical variation rather than concentrating on a single semantic field.

The questionnaire was particularly important for distinguishing geographic variation from possible age-related variation. A lexical form was not considered age-related simply because it was produced by a younger or older informant. Instead, the responses were compared across both location and age cohort. When speakers from different age groups within the same kecamatan produced the same form, while speakers in another kecamatan consistently produced a different form, the difference was interpreted primarily as geographic. Conversely, differences between age groups within the same location were examined as possible age-related variation. This procedure allowed the questionnaire data to serve as the main basis for identifying recurring lexical patterns in the study.

- Semi-structured interviews explored informants' length of residence, patterns of language use at home and work, self-reported perceptions of pronunciation and vocabulary differences between areas, attitudes toward the high speech register, and views on language preservation. Seven were transcribed and are analyzed below. The interviews were audio-recorded with a mobile phone and transcribed by the researcher. The interviews used a mixture of Indonesian and Madurese, lasted about 5–10 minutes each, and took place at the informants' homes or workplaces.

The semi-structured interviews were used to provide contextual information for interpreting the lexical forms identified in the questionnaire. The interview topics were kept sufficiently consistent across informants to allow comparison, while the semi-structured format also allowed speakers to elaborate on differences that they considered important. Particular attention was given to speakers' descriptions of where particular forms were used, their experiences of language use at home and work, their familiarity with the high speech register, and their views on language preservation. These topics were relevant to the research questions because lexical variation cannot always be interpreted from isolated forms alone. For example, a form may differ because of location, but speakers' explanations may reveal that it is also associated with politeness, social setting, or changing patterns of language use. The interview data therefore functioned as contextual evidence for interpreting patterns found in the word-list responses.

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Data Analysis

Questionnaire responses were compared across *kecamatan* and age cohorts item by item. An item was treated as geographic when the same form or set of forms recurred across all age cohorts represented within a *kecamatan* but differed between *kecamatan*; as potentially age-related when informants of different ages within the same *kecamatan* gave different forms; and as individual when a single informant produced a form without a recognizable pattern. Because there are only a few informants per cell, the comparison is descriptive: no statistical tests were applied, and counts are given only to show how many informants produced a form. Interview data were analyzed thematically (Braun & Clarke, 2006);(Miles & Huberman, 1994), focusing on informants' own explanations for lexical and register differences. The transcripts were read, and recurring information related to language use, lexical and speech-level differences, language attitudes, and language preservation was identified. Relevant statements were grouped according to these themes and compared with the questionnaire findings to identify similarities and differences between informants' reported experiences and their lexical responses. Interview statements are treated as informants' own reports and perceptions: they are used to illustrate and contextualize the word-list patterns, and, where possible, were compared against the questionnaire responses of the same individuals, but they are not counted as evidence of community-wide usage. Quotations from the interviews are given in the original Indonesian (in italics), followed by an English gloss.

FINDINGS AND DISCUSSION

FINDINGS

Geographic variation is the clearest pattern in the sample

In this sample, the clearest and most consistent variation in the word-list data runs along *kecamatan* lines rather than generational ones. The second-person pronoun *kamu* is a good example, as summarized in Table 3.

Table 3. *Kamu* "you" across *kecamatan*, all age cohorts combined

Kecamatan	Form(s) used	Consistent across the cohorts represented?
Tanjung Bumi	*kakeh*	Yes (all 4 informants)
Kokop	*kakeh* / *be'na* / *hedeh* (variants given by the same informants)	Yes, with internal variation
Labang	*hedeh*	Yes (all 3 informants)
Kota Bangkalan	*be'na* (casual) / *sampean*, *panjenengan* (polite, 1 informant)	Mostly; one register exception
Tanah Merah	*be'na*	Yes (both informants)

The pattern was also mentioned in the interviews. AM-61 (Kota Bangkalan) volunteered, without being shown the questionnaire, that *kalau di pangkalan, kamu kan Bekna. Kalau di kecamatan lain… Hede* ("in Bangkalan, 'you' is *bekna*; in other *kecamatan*… *hede*"), which matches the questionnaire data. This is one informant's observation, but it points in the same direction as the word-list responses. Similarly, the word for *pasir* "sand" splits geographically: *paser* in Tanjung Bumi and Kokop versus *beddih* in Labang, Kota Bangkalan, and Tanah Merah, in each case across the cohorts represented in the *kecamatan*. This distribution is consistent with a geographic rather than a generational pattern, although with two to four informants per *kecamatan* it should not be read as fixing a dialect boundary.

The distribution of these lexical forms suggests that geographic variation is not limited to a single vocabulary item. The differences found in the terms for "you" and "sand" show that speakers from different parts of Bangkalan may maintain different lexical choices even when they

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

refer to the same everyday concept. More importantly, the differences are not consistently associated with the age of the speakers. In Tanjung Bumi, for example, the form *kakeh* was produced by informants of different ages, while *hedeh* was consistently found among the Labang informants. This distribution makes it difficult to interpret the difference simply as a generational preference.

The same pattern can be seen in the lexical item referring to "sand." The form *paser* occurred among speakers from Tanjung Bumi and Kokop, whereas *beddih* occurred among speakers from Labang, Kota Bangkalan, and Tanah Merah. Because the forms are distributed across different locations rather than being restricted to one age group, the data provide stronger evidence for geographic patterning. These examples also show why the comparison between location and age is important in this study: a lexical form used by a younger speaker does not automatically represent language change if older speakers in the same area use the same form.

The findings therefore indicate that, in this sample, lexical variation in Bangkalan is locally distributed. The five sub-districts cannot be treated as one completely uniform speech community because speakers may share the same language while maintaining different lexical preferences according to their place of residence. This is consistent with Sugiharti and Pariyanto's (2021) finding of lexical differences among areas within Bangkalan, and with work on Javanese that reports lexical differences within a single province (Puspandari, 2026), although the present word list is much smaller than theirs, so the comparison is only indicative. At the same time, the data should not be interpreted as evidence that every *kecamatan* has a completely separate dialect. Some forms are shared across locations, while others show more localized distributions, so the variation is better understood as a pattern of overlapping lexical features than as a set of sharply separated dialect boundaries. Other explanations also remain open: with two to four informants per *kecamatan*, some forms may reflect family or village usage rather than the *kecamatan* as a whole, as one interviewee's remarks about neighboring villages suggest (see below).

A few items show possible age-related differences

Within Labang, the first-person pronoun *saya/aku* "I" differs by age: the older informant (SU-56) gave *sengko'*, while both younger informants (AR-30, HE-25) gave *engko'*. The same alternation also appears between *kecamatan* (Kota Bangkalan and Tanah Merah informants of various ages gave *sengko'*), so this item may reflect a combination of geographic and age-related conditioning. A possible case of lexical shift appears with *gigi* "teeth": fourteen of fifteen informants gave the inherited form *ghighi*, but the youngest informant in Labang (AR-30) gave *gigih*, phonologically closer to Indonesian *gigi*. This is consistent with convergence among younger speakers, though a single instance is suggestive rather than conclusive. Taken together, age-related differences in the basic word list are rare and small in this sample compared with the geographic differences.

The comparison between geographic location and age shows that the two variables cannot always be separated completely. Some lexical differences appear to involve age, but the distribution of the forms also shows that geographic location remains an important factor. The variation between *sengko*' and *engko'* provides an example. The younger speakers in Labang used *engko'*, while the older speaker used *sengko'*. At first glance, this difference may suggest a generational distinction. However, *sengko'* was also found among speakers from other locations and age groups. Therefore, the difference cannot be interpreted as a general change affecting all younger speakers in Bangkalan.

The example of *gigih* and *ghighi* should also be treated cautiously. The form *gigih* was produced by one younger speaker in Labang, while the other informants predominantly used *ghighi*. The similarity between *gigih* and the Indonesian form *gigi* makes the example relevant to possible Indonesian influence. Nevertheless, because the form occurred in only one response, it cannot be considered sufficient evidence of a broader generational change. Instead, it represents an individual or potentially emerging pattern that would require additional speakers and repeated observations to confirm.

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

These findings show the importance of distinguishing between a difference that occurs between individuals and a difference that occurs systematically between groups. An isolated form may be linguistically interesting, but it does not necessarily indicate a change in progress. A stronger age-based interpretation would require similar differences to occur repeatedly among speakers of different generations within the same location. The present dataset provides some indications of this possibility but does not provide enough evidence to establish a general pattern of generational lexical replacement.

The interaction between geography and age is therefore best described as uneven. Geographic differences are more consistently visible in the word-list data, whereas age-related differences appear in a smaller number of items. This unevenness is itself informative because it prevents the analysis from assuming that younger speakers necessarily use less traditional Madurese vocabulary than older speakers. In several cases, speakers of different ages used the same local forms, which is compatible with continued transmission of local forms across generations, although real-time evidence would be needed to show this.

Several alternative explanations should be kept in view. The *sengko'/engko'* contrast in Labang could reflect individual, family, or village usage rather than age. The form *gigih* could be an individual variant, or it could reflect the influence of Indonesian spelling or pronunciation during elicitation; this cannot be excluded on a single response. Muttaqin et al. (2019) report lexical differences related to age and politeness across regions of Madura, whereas the present word list shows few age-related differences within Bangkalan. This need not be a contradiction: the studies differ in scope (three regencies versus five sub-districts of one regency), in instrument, and in the vocabulary sampled, and the present list consists largely of basic vocabulary, which may be less sensitive to age-related change than other domains.

Limited use of the high register: a possible generational and social pattern

Of the fifteen word-list informants, only one, MZ-58 (Kota Bangkalan), spontaneously answered the entire questionnaire in the high (*èngghi-bhunten*) register, for example *asta* “hand”, *lesan* “mouth”, *kaule* “I”, *sampean/panjenengan* “you”, *rabu* “come”, and *ngoladhin* “see”. All other informants, including the two other older informants (AM-61 and SU-56), answered in the everyday register. Table 4 illustrates the contrast.

The interview data provide additional information about why the distribution of the high speech level is different from the distribution of everyday lexical forms. The use of *èngghi-bhunten* was not simply determined by age. Although the only informant who consistently produced the high register throughout the questionnaire was an older speaker, other older speakers did not necessarily demonstrate the same level of competence. This suggests that knowledge of the high register may depend on more than chronological age. Speakers may acquire and maintain the register through particular social relationships, family practices, and communicative situations.

The comments from the informants also indicate that familiarity with *èngghi-bhunten* may influence speakers’ confidence in using it. The uncertainty expressed by SI-37 and the comment from HE-25 about younger speakers feeling embarrassed or reluctant to use the high register suggest that the issue is not simply whether a speaker knows a particular vocabulary item. It may also involve whether the speaker feels comfortable using the register in interaction. In this sense, the findings connect lexical competence with language attitudes and social experience.

The limited use of *èngghi-bhunten* in the questionnaire should therefore be interpreted as an indication of reduced or uneven competence rather than as proof that the register has disappeared. The register remains recognized as part of Madurese speech, but its active use appears to be less widespread than everyday vocabulary. The fact that some older speakers also did not consistently use the high register further supports a cautious interpretation. A larger sample would be necessary to determine whether the reduced use is related to age, social background, family transmission, education, or communicative context.

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Table 4. Everyday register (14 informants) versus high register (1 informant, MZ-58)

Concept	Everyday form	High-register form (MZ-58)
Saya "I"	*engko'* / *sengko'*	*kaule*
Kamu "you"	*kakeh* / *hedeh* / *be'na*	*sampean* / *panjenengan*
Tangan "hand"	*tanang*	*asta*
Makan "eat"	*ngakan*	*adhe'er*
Datang "come"	*deteng*	*rabu*
Lihat "see"	*ajhelling*	*ngoladhin*

A single case cannot establish a generational trend, but it can be set beside the interview data. Two interviewees, SI-37 (Kokop, middle cohort) and HE-25 (Labang, younger cohort), stated that they themselves do not fully command the high register. SI-37 said the everyday register is manageable but that she is unsure about *èngghi-bhunten*: *itu yang paling tinggi itu kayaknya, saya aja pun mungkin kurang ngerti ya kalau sekarang* ("that's the highest level; even I'm probably not very sure about it these days"). HE-25 linked this directly to youth attitudes: *saya sendiri aja kurang paham kalau bahasa halusnya yang tingkatan tinggi… itu mungkin yang bikin malu atau bikin anak-anak muda itu males gunain bahasa Madura* ("I myself don't really understand the high polite level… that might be what makes young people embarrassed or reluctant to use Madurese").

The rarity of the high register in the word-list data and these two interviewees' remarks point in the same direction, but the evidence is thin. In the questionnaire, the everyday register was used by two of the three older informants and by all six middle-aged and all six younger informants, and one middle-aged interviewee (SI-37) described her own uncertainty. This pattern is compatible with reduced or uneven command of the high register beyond the youngest cohort, but it does not show a decline. The questionnaire recorded one register per informant, so an informant who answered in the everyday register may nevertheless know the high-register forms, and register choice may also reflect the elicitation situation (for example, how the informant viewed the researcher) and not competence alone.

Individual reports of lexical loss and borrowing

- DH-33 (Tanah Merah), a schoolteacher, reported that her own child no longer recognizes a basic body-part term: *Kadang dia, "Mi, apa itu obu'?" Rambut itu apa, Mi? Kalau dulu jaman saya sekolah itu obuh* ("Sometimes she asks, 'Mom, what's *obuh*?' What's the word for hair, Mom? Back when I was in school, it was *obuh*"). This is a first-hand report of a possible case of lexical loss in one child's vocabulary. It comes from one parent about one child and cannot show that *obuh* is being lost more widely.
- AM-61 (Kota Bangkalan) described the replacement of inherited colour terms by loanwords: *Kalau dulu merah ngode. Kalau sekarang bukan merah ngode, pink sekarang… Kalau dulu orange… Konyek kapor* ("In the past, [that color] was called *merah ngode*; now it's just 'pink'. In the past, [orange] was *konyek kapor*; now it's 'orange'"). This is an older speaker's report about his own current usage, so in this account borrowing is not only a youth phenomenon; but it is a single report and needs confirmation from other speakers.
- SI-37 (Kokop) described village-level, not just *kecamatan*-level, variation: *Jangankan ke luar kecamatan, desa ada… dari sini ke Desa sebelah. Itu ada dialek-dialek seperti itu* ("Even without leaving the *kecamatan*, there are differences between villages, for instance between here and the next village"), citing *apapoh* "what" as used in a neighboring village but not her own. The relevant boundaries may thus be finer-grained than the *kecamatan* level assumed in the research design.

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Self-reported Madurese use across communities

Across interviews, informants' self-estimated proportion of Madurese (versus Indonesian) use appeared to track proximity to the regency's urban center more closely than it tracked age, but this impression rests on a handful of estimates. SE-20 (Kokop) estimated 90–95% Madurese use; SI-37 (Kokop) estimated 70%, contrasting this with Bangkalan, Burneh, and Tanah Merah, which she described as *udah minim* ("already minimal"); AM-61 (Kota Bangkalan) estimated a 50/50 split. MU-31 (Tanjung Bumi), from a coastal fishing community, reported that fishers *tetap berkomunikasi dengan Bahasa Madura* ("still communicate in Madurese") and described Indonesian use as *minim* ("minimal"). Although she belongs to the younger cohort, she reports a Madurese-dominant environment, which suggests that younger speakers do not uniformly use less Madurese. These figures are self-estimates, not measurements of actual usage, and they come from individuals who differ in occupation, education, and household situation, so place cannot be separated from these other factors. They are reported as perceptions of the language ecology that merit further investigation, not as an established gradient.

Institutional and attitudinal context

Several informants referred to the local-language subject (*Bahasa Daerah*) in schools as having been discontinued for a period and only recently reinstated, and none of the seven interviewees knew of any government preservation program outside the school curriculum, with the partial exception of a school-based "Kacong Cebhing" ambassador program mentioned by HE-25. Informants used the words *minder* (self-conscious/inferior) and *malu* (ashamed) to explain why younger speakers in particular avoid Madurese or suppress their Madurese accent outside Madura, while expressing personal disagreement with erasing the accent. DH-33 said: *kalau untuk pelestarian, saya tidak setuju kalau dihilangkan* ("as far as preservation goes, I don't agree with getting rid of it"). These statements are informants' perceptions of other speakers' behavior and were not checked against observation of actual language choice.

Summary: observed patterns and interpretations

Table 5 brings together the main patterns and indicates how each should be interpreted given the current sample.

Table 5. Summary of the main patterns in the data

Pattern	Evidence	Interpretation
Geographic variation	*Kamu* (*kakeh*, *hedeh*, *be'na*); *pasir* (*paser* vs. *beddih*)	Clearest pattern in the sample; consistent with sub-district-level lexical differences, but based on two to four informants per *kecamatan*
Possible age effect	*Sengko'* vs. *engko'* in Labang; *gigih* vs. *ghighi*	Suggestive only; confounded with place or based on one speaker
Limited use of the high register	One of fifteen informants used the high register; two interviewees report limited command	Possibly reduced or uneven command; cannot be attributed to age, social background, or the elicitation situation without a larger sample
Lexical loss and borrowing	*Obuh* (one parent's report); colour terms replaced by *pink* and *orange* (one older speaker)	Individual reports; possible contact-induced change that needs confirmation
Self-reported Madurese use	Self-estimates of Madurese use: 70–95% in Kokop vs. about 50% in Kota Bangkalan	Perceived language ecology; place, occupation, and individual factors cannot be separated

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

DISCUSSION

The findings can be related to previous work in four ways, keeping observed patterns separate from interpretation. First, the predominance of geographic variation in the sample is consistent with the dialectological expectation that everyday vocabulary is distributed in regional patterns (Chambers & Trudgill, 1998) and with the lexical differences among areas that Sugiharti and Pariyanto (2021) report for Bangkalan. What was observed is that *kamu* and *pasir* differ by *kecamatan* in the same way in all the cohorts represented. As an interpretation, such items can be classified as stable rather than changing in an apparent-time analysis, because no age gradient suggests change in progress (Tagliamonte, 2012). One informant's unprompted description of the same distribution suggests that speakers recognize at least some of these differences, but a single remark does not show that the boundaries form part of the community's shared metalinguistic awareness. Alternative explanations include family or village-level usage and the small number of informants per *kecamatan*. The result also parallels studies of other Indonesian languages that find regional lexical differences within one province (Puspandari, 2026).

Second, the register data show that apparent-time reasoning must be applied with care in a speech-level language. What was observed is that only one informant (MZ-58, older) answered in the high register, that other older and middle-aged informants did not, and that two interviewees said they do not fully command it. One possible interpretation is that command of the high register is receding across a broad span of the adult population, which would parallel reports of weakening formal speech styles among younger Javanese speakers (Errington, 1998); (Smith-Hefner, 2009) and of reduced use of speech levels among young Balinese (Sutika, 2019). Alternative explanations are at least as plausible given the data. Register choice may reflect the elicitation situation, the informant's relationship with the researcher, occupation, or family background rather than competence, and age grading rather than generational change cannot be excluded without real-time data (Holmes & Wilson, 2017). In languages such as Madurese, a form is selected according to the social relationship between speaker and addressee (Stevens, 1968), so a speaker may understand a polite form and yet avoid it because the situations that call for it have become rarer or because using it feels unnatural. The reluctance that some informants described (*malu* and *minder*) points to an attitudinal component that a word list alone cannot capture, but these remarks come from a few interviewees and cannot establish a community-wide attitude. Because the register is learned through use in formal and hierarchical settings (Holmes & Wilson, 2017), any loss would be closely tied to changes in the settings in which it is practiced; this remains to be tested.

Third, the individual reports of lexical loss and borrowing are compatible with the contact-linguistic view that the lexicon is the most permeable part of a language (Thomason & Kaufman, 1988). The replacement of *merah ngode* and *konyek kapor* by *pink* and *orange*, as reported by one older speaker, would fit a pattern in which borrowing proceeds in domains where foreign terms are widely available through media and education. The reported loss of *obuh* would fit the possibility that even basic vocabulary can be affected when Indonesian, which has the equivalent *rambut*, becomes the main source of input for children. These are interpretations of single reports, however, and other explanations are possible: the child may simply not have encountered the word, or the word may be restricted to certain families or villages, and colour terms are a domain in which borrowing is common without implying wider loss. The reports therefore indicate where pressure from Indonesian might be visible and where further data should be collected (Fishman, 1991). They do not show that core vocabulary is being lost or that Madurese is in danger in Bangkalan, particularly given the high levels of use reported in Kokop and Tanjung Bumi.

Finally, the village-level differences and the differences in self-reported use between communities suggest that age is not the only social dimension that may matter. Place of residence and occupation may shape exposure to Indonesian and therefore the extent to which younger speakers maintain Madurese, but the present data cannot separate these factors from individual variation. A fuller sociolinguistic account of Bangkalan would need to combine age, place, and occupation in a larger and more balanced sample, and a *kecamatan*-based design can only approximate spatial detail.

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

These findings carry tentative practical implications. Because everyday vocabulary varied by locality in this sample, teaching materials for Madurese in Bangkalan could acknowledge that the same concept may be expressed by different words in different *kecamatan*, and teachers could treat local forms as legitimate rather than as deviations from a single norm. Because the high register appeared in only one informant's questionnaire and some interviewees reported limited command, schools and community organizations could consider giving explicit attention to the speech levels and the situations in which they are used, so that younger speakers can learn them in context. Because informants mentioned embarrassment as a reason for avoiding Madurese, maintenance efforts could also address the social image of Madurese (Fishman, 1991); informants' statements that they do not wish to see the accent disappear indicate goodwill on which such efforts can build. Finally, since none of the seven interviewees knew of preservation programs beyond the school curriculum, better communication about existing initiatives, such as the school-based ambassador program, could increase their reach.

CONCLUSIONS AND SUGGESTIONS

Based on 15 word-list informants and 7 interview transcripts across Kota Bangkalan, Labang, Tanah Merah, Kokop, and Tanjung Bumi, geographic variation is the clearest pattern in the sample. Forms such as those for *kamu* and *pasir* differed by *kecamatan* in the age cohorts represented, and interviewees also mentioned village-level differences, whereas age-related differences in the word list were few and, where they occurred, either could not be separated from place or rested on a single speaker. Command of the high (*èngghi-bhunten*) speech register is the one domain in which a possible generational or social pattern is suggested: it was nearly absent from the word-list data and two interviewees reported limited command. This pattern requires confirmation with a larger sample and cannot yet be separated from age grading, elicitation effects, or individual background. The interviews additionally contained individual reports of lexical loss (*obuh* "hair") and borrowing (colour terms) and suggested that self-reported Madurese use may be higher in Kokop and Tanjung Bumi than in the regency capital; these are exploratory observations. Age-related differences and lexical change therefore require further investigation, and the study does not show that Madurese is declining or that intergenerational transmission is weakening in Bangkalan.

Several suggestions follow. For documentation, descriptions of Bangkalan Madurese should record variation at the level of *kecamatan* and, where possible, village rather than treat the regency as uniform. For language maintenance, school-based instruction and community activities could give attention to polite speech levels and to everyday core vocabulary, not only to cultural performance. For future research, extending the word list to fill the remaining age-by-*kecamatan* gaps, increasing the number of informants per cell, completing the remaining transcriptions, recording elicitation sessions, and revisiting the same communities in real time would allow the patterns reported here to be tested more rigorously.

LIMITATIONS

The current sample (15 word-list informants, unevenly distributed across age cohorts and *kecamatan*; 7 of a larger planned set of interviews transcribed) means the findings should be read as preliminary. Each *kecamatan* has only two to four informants and none has informants from all three cohorts, so geographic and age comparisons are partly confounded and the patterns cannot be generalized to a *kecamatan* or to Bangkalan as a whole. No older-generation word-list data are available for Tanjung Bumi, Kokop, or Tanah Merah, no middle-generation data for Labang, and no younger-generation data for Kota Bangkalan. Several findings rest on individual responses (*gigih*; the high-register answers of MZ-58) or on interviewees' personal observations (*obuh*, colour terms, self-estimated language use). These are useful exploratory findings, but they cannot establish community-wide patterns. The apparent rarity of the high register rests on a single clear case and should be confirmed with a larger sample before it is treated as a generational pattern rather than individual, occupational, or situational variation. In addition, the apparent-time

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

interpretation cannot distinguish generational change from age grading without real-time data, and the self-reported figures for Madurese use are estimates that may not match actual usage.

REFERENCES

Afiansyah, R. M. (2023). Ondhâgga Bâsa, Levels of Speech in Madurese: Sociolinguistic Studies in Language Variations. *Linguistik, Terjemahan, Sastra (LINGTERSA)*, *4*(2), 62–68. https://doi.org/10.32734/lingtersa.v4i2.10765

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chambers, J. K., & Trudgill, P. (1998). *Dialectology* (2nd ed.). Cambridge University Press.

Davies, W. D. (2010). *A Grammar of Madurese*. De Gruyter Mouton. https://doi.org/10.1515/9783110224443

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2023). *Ethnologue: Languages of the World* (26th ed.). SIL International. https://www.ethnologue.com

Errington, J. (1998). *Shifting Languages: Interaction and Identity in Javanese Indonesia*. Cambridge University Press.

Fishman, J. A. (1991). *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Multilingual Matters. https://doi.org/10.21832/9781800418097

Holmes, J., & Wilson, N. (2017). *An Introduction to Sociolinguistics* (5th ed.). Routledge.

Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). SAGE Publications.

Muttaqin, S., Sahiruddin, & Rodliyah, I. N. (2019). Language Variations in Madurese across Regions and Age Groups: Looking at Syntactic and Lexical Variations among Regions and Age Groups. *KLAUSA: Kajian Linguistik, Pembelajaran Bahasa, Dan Sastra*, *3*(1), 45–56. https://doi.org/10.33479/klausa.v3i01.193

Puspandari, L. E. (2026). Dialectological Analysis of East Java's Arekan, Mataraman, and Pendhalungan Regional Linguistic Features. *Indonesian Journal of Applied Linguistics*, *15*(3), 658–672. https://doi.org/10.17509/6kd1sp97

Smith-Hefner, N. J. (2009). Language Shift, Gender, and Ideologies of Modernity in Central Java, Indonesia. *Journal of Linguistic Anthropology*, *19*(1), 57–77. https://doi.org/10.1111/j.1548-1395.2009.01019.x

Sneddon, J. N. (2003). *Indonesian: A Comprehensive Grammar* (1st ed.). Routledge.

Stevens, A. M. (1968). Madurese Phonology and Morphology. In *American Oriental Series* (Vol. 52). American Oriental Society.

Sugiharti, D. R., & Pariyanto. (2021). Lexical Variation of Madurese Spoken in Bangkalan, Madura. *PARAFRASE: Jurnal Kajian Kebahasaan & Kesastraan*, *21*(1). https://doi.org/10.30996/parafrase.v21i1.5253

Sutika, I. N. D. (2019). Penggunaan Anggah Ungguhing Basa Bali: Sebuah Kesantunan dan Penanda Kelas Sosial Masyarakat Bali. *Pustaka: Jurnal Ilmu-Ilmu Budaya*, *19*(2), 68–73. https://doi.org/10.24843/PJIIB.2019.v19.i02.p02

Tagliamonte, S. A. (2012). *Analysing Sociolinguistic Variation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511980425

Thomason, S. G., & Kaufman, T. (1988). *Language Contact, Creolization, and Genetic Linguistics*. University of California Press. https://doi.org/10.1525/9780520912793

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical Foundations for a Theory of Language Change. In *Directions for Historical Linguistics*. University of Texas Press.

Copyright (c) 2026 Khairatul Amaliya, Zulkaeriyah, Diah Ekawati Ayuningtias, Desi Puspitasari, Agustinus GR Dasion. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.