

RELIGIOUS CONTACT AND LOANWORD STRATIFICATION IN INDONESIAN: A CORPUS-BASED STUDY

Jean Paul Gotopo

Universidad Pedagógica Experimental Libertador, Maracay, Venezuela

E-mail: jean.gotopo.ipmar@upel.edu.ve

Received: 2026-02-21

Accepted: 2026-03-26

Published: 2026-06-08

Abstract

This study examines how successive waves of religious contact—including Hindu-Buddhist, Islamic, and Christian influences—have contributed to loanword stratification in the Indonesian language. Religious contact has been a major driver of lexical borrowing across multilingual societies, yet its role in producing stratified loanword systems within national languages remains insufficiently understood. However, few studies have examined how successive waves of religious contact produce stratified lexical systems in contemporary Indonesian. Using a corpus-based analysis of 1,402 lexical items compiled from major dictionaries and national linguistic corpora, the study classifies loanwords according to their etymological origins and semantic domains. The results reveal a layered pattern of borrowing aligned with historical phases of religious and cultural contact. Sanskrit-derived terms constitute the largest group (31.2%), followed by Dutch (24.2%), Arabic (13.1%), and Portuguese (4.6%), with smaller contributions from Hindi, Pali, and regional languages. While these lexical strata reflect distinct historical influences, many borrowed forms have become fully integrated into contemporary Indonesian across religious, institutional, and everyday domains. The findings demonstrate that religious contact contributed not only to specialized religious vocabulary but also to broader layers of lexical borrowing that now form part of a shared sociolinguistic repertoire. By linking historical language contact with patterns of lexical integration, this study highlights how religious interaction can produce enduring stratification within the Indonesian lexicon.

Keywords: *Corpus-based analysis; Indonesian language; lexical borrowing; lexical stratification; religious contact; sociolinguistics.*

1. Introduction

Religious contact has long been recognized as an important driver of linguistic change in multilingual societies. Throughout history, the spread of religious traditions has facilitated the movement of lexical items across linguistic boundaries through trade, missionary activity, education, and institutional governance (Nesher, 2025, p. 40). These processes often leave lasting traces in national vocabularies, particularly in semantic domains related to ritual practice, ethics, law, and social organization. Despite the acknowledged importance

of religious contact in language history, the mechanisms through which successive religious encounters produce stratified lexical systems in contemporary languages remain theoretically underdeveloped and empirically underexplored.

Previous studies on language contact in Southeast Asia have documented extensive lexical borrowing from Sanskrit, Arabic, and various European languages into Malay and Indonesian (De Casparis et al., 1997, p. 11; Julul et al., 2020, p. 143). Research has examined phonological adaptation, semantic change, and the historical diffusion of loanwords within the region. However, much of this scholarship focuses on individual source languages or on etymological cataloguing of borrowed vocabulary. Fewer studies have examined how multiple waves of religious and cultural contact interact over time to generate layered lexical systems that remain active in modern language use (Harahap et al., 2025, p. 29).

Indonesia provides a particularly revealing context for examining this phenomenon. As the world's largest Muslim-majority nation and one of the most linguistically diverse societies in the world, Indonesia has experienced successive periods of religious and cultural contact over more than a millennium. Hindu-Buddhist kingdoms, the spread of Islam through maritime trade networks, and the arrival of Christian missions during the colonial period each contributed new lexical resources to the evolving Malay-Indonesian linguistic system. As a result, contemporary Indonesian contains a complex mixture of loanwords originating from Sanskrit, Arabic, Portuguese, Dutch, and other regional languages, reflecting different historical phases of religious interaction (Musgrave, 2014, p. 89; Ricklefs, 1993, p. 8).

These historical encounters have left visible traces in everyday language use. Many lexical items originally associated with specific religious traditions now appear across a wide range of social contexts, including administration, education, media discourse, and daily conversation. Expressions such as *Insha Allah*, *Alhamdulillah*, and *Amin*—derived from Arabic—are widely used in everyday interaction, while Sanskrit-derived terms such as *agama* or *bahasa* function as neutral institutional vocabulary (Abdulkarim et al., 2020, p. 378). This coexistence suggests that religiously mediated borrowing may contribute not only to lexical expansion but also to the formation of shared linguistic repertoires across communities.

This study investigates how religious contact has contributed to the formation and lexical stratification of loanwords in the Indonesian language. Drawing on a corpus-based analysis of 1,402 lexical items compiled from major dictionaries and national linguistic corpora, the study examines the etymological origins of religiously associated vocabulary and explores how these lexical layers correspond to historical periods of religious influence.

Specifically, this study addresses three main research questions:

1. What proportion of Indonesian vocabulary derives from religiously associated languages?
2. Through which historical processes did these loanwords enter Indonesian?
3. How are these loanwords used in contemporary Indonesian?

By linking historical etymology with patterns of contemporary language use, this study contributes to broader discussions in sociolinguistics and historical linguistics concerning the role of cultural and religious contact in shaping national languages. The findings highlight how successive waves of religious interaction have produced a stratified

lexical system in Indonesian that continues to function as part of a shared sociolinguistic repertoire in a multilingual society.

2. Literature Review

The Indonesian lexicon reflects multiple layers of linguistic contact shaped by successive religious and cultural encounters. Previous research has extensively documented lexical borrowing from Sanskrit, Arabic, and European languages, highlighting the role of trade, religion, and colonialism in shaping the vocabulary of Malay and Indonesian (Mahfud et al., 2021, p. 229; Rokib et al., 2023, p. 530). These studies have demonstrated how borrowed terms undergo phonological adaptation and semantic change, becoming integrated into local linguistic systems (Batais & Wiltshire, 2022, p. 247; Julul et al., 2020, p. 139).

However, much of the existing literature has focused on individual source languages or on descriptive cataloguing of loanwords. Research on Sanskrit influence emphasizes its role in governance, philosophy, and literary traditions, while studies on Arabic borrowing highlight its importance in religious discourse and everyday communication (Alnizar et al., 2025). Similarly, scholarship on European influence, particularly Portuguese and Dutch, has examined the introduction of administrative and institutional vocabulary during the colonial period.

While these studies provide valuable insights into the origins and adaptation of loanwords, fewer have examined how these multiple layers of borrowing interact to form a stratified lexical system within contemporary Indonesian. In particular, the relationship between religious contact and the long-term structuring of the lexicon remains underexplored. Existing research rarely considers how successive waves of religious influence contribute not only distinct lexical items but also enduring patterns of coexistence among different etymological layers.

Recent studies in sociolinguistics and language contact theory suggest that borrowing processes in multilingual societies often operate through accumulation rather than replacement, resulting in layered lexical systems that reflect historical trajectories of cultural interaction (Harahap et al., 2025, p. 30). From this perspective, the Indonesian lexicon can be understood as a product of cumulative contact, in which Sanskrit, Arabic, and European elements coexist and interact within a shared linguistic repertoire.

Building on this perspective, the present study examines how loanwords introduced through religious contact contribute to lexical stratification in Indonesian. By combining corpus-based analysis with historical interpretation, this study seeks to bridge the gap between etymological description and sociolinguistic explanation, providing a more integrated account of how religious contact shapes the structure of the lexicon over time.

2.1 Sanskrit Foundations

The earliest layer of lexical borrowing in Indonesian derives from Sanskrit, introduced through Hindu-Buddhist cultural and political influence in the Indonesian archipelago. Previous studies have shown that Sanskrit functioned as a language of governance, religion, and intellectual discourse, contributing a substantial number of lexical items to Old Malay and later Indonesian (Gil, 2024, p. 490).

Sanskrit-derived loanwords are particularly prominent in domains such as governance (*raja*, *desa*), philosophy (*jiwa*, *rasa*), and law (*jaksa*) (Mahfud et al., 2021). These terms illustrate how early cultural and religious contact established a foundational layer of vocabulary that continues to shape the Indonesian lexicon.

Importantly, this Sanskrit layer has not been replaced by later borrowings but has remained embedded within the language, coexisting with subsequent lexical influences. This pattern supports the view that lexical borrowing in Indonesian operates through cumulative layering rather than replacement, forming a stratified lexical system over time.

2.2 Arabic Enrichment

The spread of Islam introduced a substantial layer of Arabic-derived vocabulary into Malay and Indonesian, primarily through trade networks, religious education, and institutional practices. Previous research has documented the extensive influence of Arabic on the Indonesian lexicon, with loanwords spanning both religious and everyday domains (Alnizar et al., 2025; Julul et al., 2020, p. 139).

Arabic loanwords are particularly prominent in religious discourse, including terms such as *iman* ("faith"), *masjid* ("mosque"), and *zakat* ("almsgiving"). However, their influence extends beyond strictly theological contexts, with many items becoming integrated into everyday language, such as *dunia* ("world") and *kursi* ("chair"). This dual distribution illustrates how lexical items introduced through religious contact can expand into broader semantic domains.

Studies have also shown that Arabic borrowings in Indonesian often undergo semantic specialization or extension. For example, *shalat* retains a specific religious meaning, while *ayat* has expanded to include broader interpretations such as "legal clause" or "sign" (Alnizar et al., 2025). These processes highlight the dynamic adaptation of borrowed vocabulary within local linguistic systems.

Importantly, the integration of Arabic loanwords reflects not only linguistic processes but also sociocultural dynamics. As noted in previous studies, the codification of Arabic-derived vocabulary in official dictionaries such as the *Kamus Besar Bahasa Indonesia* (KBBI) can influence which forms are standardized and widely adopted (Rokib et al., 2023, p. 531). This suggests that lexical borrowing is shaped not only by contact but also by institutional and ideological factors.

From the perspective of lexical stratification, Arabic-derived vocabulary constitutes a distinct but overlapping layer within the Indonesian lexicon. Rather than replacing earlier Sanskrit-derived terms, Arabic loanwords coexist with them, contributing to a cumulative and stratified lexical system shaped by successive waves of religious contact.

2.3 European Layers

European contact introduced an additional layer of lexical borrowing into Indonesian, primarily through Portuguese and Dutch influence during the colonial period. These borrowings differ from earlier Sanskrit and Arabic layers in that they are less directly tied to religious transmission and more closely associated with administration, education, and technological change (Wijaya et al., 2021, p. 90).

Portuguese influence, dating from the sixteenth century, contributed a limited but significant set of loanwords, many of which are associated with religion, trade, and material culture. Examples include *gereja* ("church"), *Natal* ("Christmas"), and *sabun* ("soap").

Although smaller in scale compared to later borrowings, Portuguese lexical items represent an early stage of European linguistic influence in the region.

Dutch influence, by contrast, constitutes one of the largest layers of borrowing in Indonesian. During several centuries of colonial rule, Dutch served as the language of administration, education, and governance, introducing a wide range of lexical items into Malay and later Indonesian. These include terms such as *kantor* (“office”), *sekolah* (“school”), and *universitas* (“university”), which remain central to modern institutional vocabulary.

From a sociolinguistic perspective, European loanwords illustrate how colonial and institutional power structures shape lexical adoption and standardization (Harahap et al., 2025, p. 34). Unlike earlier borrowings associated with religious transmission, Dutch-derived vocabulary became embedded in formal and bureaucratic domains, reflecting the role of colonial governance in structuring linguistic change.

In terms of lexical stratification, European borrowings constitute a later layer that coexists with earlier Sanskrit and Arabic elements. Rather than displacing these earlier forms, European-derived terms expand the functional range of the lexicon, contributing to a cumulative and multi-layered linguistic system shaped by successive historical encounters.

2.4 Mechanisms of Adaptation

Lexical borrowing in Indonesian involves systematic processes of phonological, morphological, and semantic adaptation that enable foreign lexical items to be integrated into the local linguistic system. Previous studies have shown that borrowed forms are typically adjusted to conform to Indonesian phonotactic constraints, often resulting in simplified consonant clusters and vowel substitutions (Batais & Wiltshire, 2022; Julul et al., 2020, p. 138).

Phonological adaptation is particularly evident in borrowings from European languages. For example, Dutch *schroef* becomes *sekrup* in Indonesian, illustrating the restructuring of consonant clusters to fit Indonesian phonological patterns (Ricklefs, 1993, p. 11). Similarly, Arabic loanwords may undergo phonetic simplification while retaining recognizable morphological structures, as seen in forms such as *zakat* and *iman*.

In addition to phonological change, semantic adaptation plays a central role in the integration of loanwords. Borrowed terms may undergo specialization or extension in meaning depending on their usage context. For instance, certain religious terms retain their original meanings, while others expand into broader domains of everyday usage. These processes reflect the dynamic interaction between imported lexical forms and local semantic systems (Julul et al., 2020, p. 138).

Morphological integration further contributes to the assimilation of loanwords, as borrowed items become compatible with Indonesian affixation patterns. This allows loanwords to participate fully in grammatical processes, reinforcing their incorporation into the lexicon.

From a broader perspective, these mechanisms of adaptation facilitate the long-term stability of borrowed vocabulary within Indonesian. By enabling lexical items from different historical sources to conform to a shared linguistic system, phonological and semantic adaptation support the coexistence of multiple etymological layers. This process is central to the formation of a stratified lexicon shaped by successive waves of language contact.

2.5 Sociolinguistic Functions of Loanwords in Indonesian

Loanwords in Indonesian do not merely reflect historical processes of language contact; they also perform important sociolinguistic functions in contemporary communication. Rather than being restricted to specific domains, borrowed lexical items operate as markers of register, identity, and social interaction, shaping how speakers position themselves within diverse communicative contexts (Lauder, 2010).

One key function of loanwords is their role in signaling formality and institutional authority. Sanskrit- and Dutch-derived terms are frequently associated with formal registers, particularly in administrative, legal, and educational contexts. Their continued use reflects the historical prestige of these linguistic sources and their institutionalization within official discourse. In contrast, Arabic-derived expressions often function in both formal and informal settings, bridging religious and everyday communication.

Loanwords also contribute to the expression of social identity and cultural affiliation. The use of specific lexical items can index religious belonging, educational background, or participation in particular speech communities (Alsaawi, 2020, p. 249). For example, the incorporation of Arabic expressions in everyday interaction may signal not only religious orientation but also shared norms of politeness and sincerity. In this sense, borrowed vocabulary operates as a resource for constructing social meaning rather than merely conveying lexical content.

From a sociolinguistic perspective, the coexistence of multiple lexical layers enables speakers to navigate different communicative contexts by selecting from a stratified repertoire. This flexibility allows Indonesian speakers to shift between registers and social meanings, drawing on Sanskrit-derived formal vocabulary, Arabic-derived expressions of religiosity, and European-derived institutional terminology. Such patterns illustrate how lexical stratification is not only a historical phenomenon but also an active component of contemporary language use.

These observations suggest that loanword stratification in Indonesian is sustained not only by historical processes of contact but also by ongoing sociolinguistic practices. The persistence of multiple lexical layers reflects the ways in which speakers continuously reinterpret and deploy borrowed forms to negotiate identity, authority, and social relationships within a multilingual and culturally diverse society (Julul et al., 2020).

2.6 Research Gaps and Future Directions in Loanword Stratification

Despite extensive research on lexical borrowing in Indonesian, several important gaps remain in the literature. Existing studies have largely focused on identifying the origins of loanwords and describing processes of phonological and semantic adaptation (Julul et al., 2020, p. 139; Mahfud et al., 2021, p. 230; Rokib et al., 2023, p. 528). While these contributions have significantly advanced understanding of language contact in the region, they tend to examine borrowing from individual source languages in isolation, without fully accounting for the interaction between multiple layers of lexical influence.

One key gap concerns the lack of integrative approaches that analyze how successive waves of religious and cultural contact contribute to the formation of a stratified lexicon. Previous research has documented Sanskrit, Arabic, and European influences separately, but fewer studies have explored how these layers coexist and interact within contemporary Indonesian (Alnizar et al., 2025). As a result, the cumulative and structured nature of lexical borrowing remains insufficiently theorized.

A second gap relates to the limited attention given to the relationship between historical borrowing processes and present-day language use. While many studies emphasize etymology, fewer examine how loanwords continue to function across different sociolinguistic contexts. In particular, the ways in which speakers draw on multiple lexical layers to construct meaning, negotiate identity, and navigate communicative settings remain underexplored.

Additionally, methodological limitations persist in the study of Indonesian loanwords. Much of the existing research relies on relatively small datasets or qualitative examples, which may not capture broader patterns of lexical distribution (Rokib et al., 2023, p. 530). There is a need for corpus-based approaches that allow for systematic comparison across different etymological sources and semantic domains, enabling a more comprehensive understanding of lexical stratification.

Future research should therefore adopt more integrative and data-driven frameworks that combine historical, sociolinguistic, and corpus-based perspectives. Such approaches would make it possible to examine not only the origins of loanwords but also their distribution, functional differentiation, and long-term stability within the lexicon. In this regard, the present study contributes to addressing these gaps by providing a corpus-based analysis of loanword stratification in Indonesian, linking historical processes of religious contact with patterns of contemporary language use.

2.7 Historical and Religious Background

2.7.1 Hinduism and the Sanskrit Tradition

The earliest major religious influence on the Indonesian archipelago was Hinduism, which arrived through trade networks linking South Asia and Southeast Asia during the early centuries of the Common Era. Hindu-Buddhist kingdoms such as Srivijaya and Majapahit played an important role in disseminating Sanskrit as a language of religion, administration, and intellectual culture (Saputra et al., 2024, p. 250). Through these institutions, a substantial number of Sanskrit lexical items entered Old Malay and later Indonesian.

Many of these borrowings relate to governance, philosophy, and social organization. Examples include *agama* ("religion"), *bahasa* ("language"), *sastra* ("literature"), *raja* ("king"), *putra* ("son"), *putri* ("daughter"), and *manusia* ("human"). While some words retain meanings close to their Sanskrit origins, others have undergone semantic change over time. For instance, *svarga* evolved into *surga* ("heaven"), and *guru*, originally referring to a spiritual teacher, now also denotes a general instructor (Gu, 2025, p. 16).

Sanskrit loanwords frequently entered Indonesian through intermediary languages such as Old Javanese, which preserved Sanskrit vocabulary in literary traditions including kakawin poetry and court chronicles. The prominence of Sanskrit-derived vocabulary in the Indonesian lexicon reflects the long-standing cultural prestige of Hindu-Buddhist civilization in the region (De Casparis et al., 1997, p. 13).

2.7.2 Buddhism and Pali Influence

Buddhism spread to the Indonesian archipelago through the same maritime trade networks that facilitated Hindu influence. By the seventh century, the Srivijaya kingdom in Sumatra had become a significant center of Buddhist learning, attracting scholars and

pilgrims from across Asia. Early Buddhist texts in the region were written primarily in Sanskrit, although the *Theravāda* tradition later introduced elements of Pali vocabulary (Makin, 2016, p. 13).

Pali-derived loanwords in Indonesian are fewer than Sanskrit borrowings but remain significant in religious discourse. Terms such as *bhikkhu* (“monk”), *dana* (“charity”), *dukkha* (“suffering”), *karma* (“action”), and *nirwana* (“liberation”) reflect the doctrinal vocabulary of Buddhist traditions. These borrowings illustrate how religious concepts circulated through regional intellectual networks and became incorporated into the evolving Malay-Indonesian lexicon (Manguin, 2021, p. 89).

The coexistence of Hindu and Buddhist traditions also fostered a syncretic cultural environment in which Sanskrit and Pali vocabulary circulated widely. As a result, many religious and philosophical terms entered the linguistic repertoire of the region before the arrival of later religious traditions (Gu, 2025, p. 26).

2.7.3 Islam and Arabic Influence

Islam reached the Indonesian archipelago between the thirteenth and fifteenth centuries through maritime trade networks connecting the region with the Middle East, India, and the wider Muslim world. As Islam spread, Arabic became an important source of religious vocabulary in Malay and Indonesian, particularly through Qur’anic recitation, religious education, and legal discourse (Mahfud et al., 2021).

Arabic loanwords in Indonesian include core theological terms such as *iman* (“faith”), *Islam*, *Muslim*, *ulama* (“religious scholar”), *imam* (“prayer leader”), *syariah* (“Islamic law”), and *zakat* (“almsgiving”). In addition to explicitly religious vocabulary, many Arabic-derived words have entered broader domains of everyday language, including *kabar* (“news”), *daftar* (“list”), and *jadwal* (“schedule”) (Ricklefs, 1993, p. 1).

Arabic expressions are also widely used in daily interaction. Greetings and formulaic expressions such as *Assalamualaikum*, *Insyallah*, and *Alhamdulillah* function not only in religious contexts but also as markers of politeness and social identity. The widespread use of such expressions illustrates how religiously derived vocabulary can extend beyond strictly theological domains and become integrated into everyday discourse (Suri et al., 2020, p. 33).

2.7.4 Christianity and European Influence

The arrival of Europeans in the sixteenth century introduced Christianity and new languages to the archipelago. Portuguese traders captured Malacca in 1511 and established outposts in the Moluccas (Musgrave, 2014, p. 88). Along with goods and missionaries, they brought the Catholic faith and the Portuguese language. Portuguese words found their way into Malay and Indonesian via the *papiá cristão* contact language used in trade and evangelization. Religious Portuguese loanwords include *gereja* (“church”, from *igreja*), *baptis* (“baptism”, from *batismo*), *Natal* (“Christmas”), *Kristus* (“Christ”), *Paulus* (“Paul”) and *Maria* (“Mary”). Many other Portuguese borrowings are secular: *bola* (“ball”), *boneka* (“doll”), *bendera* (“flag”), *kereta* (“wagon”), *mentega* (“butter”), *almari* (“cupboard”), *tempo* (“time”), *meja* (“table”), *jendela* (“window”), *sepatu* (“shoe”), *garpu* (“fork”) and *keju* (“cheese”). These words reflect the material culture and technology introduced by the Portuguese.

Dutch involvement in Indonesia began with the establishment of the Dutch East India Company (VOC) in 1602. After displacing the Portuguese, the Dutch gradually consolidated control over trade and territory, culminating in the colonial administration of the Dutch East Indies (Braithwaite et al., 2010). The Dutch language dominated education, law, administration and commerce for more than three centuries. Consequently, Dutch has bequeathed a large number of loanwords to Indonesian—340 items (24 percent) in our corpus. Religious terms include *gereformeerd* (“Reformed”) and *dominee* (“minister”). Many more borrowings are secular: *kantor* (“office”), *asuransi* (“insurance”), *dompet* (“wallet”), *kulkas* (“refrigerator”), *apel* (“apple”), *kursus* (“course”), *stasiun* (“station”), *bioskop* (“cinema”), *polisi* (“police”), *sinterklas* (“Santa Claus”), *negara* (“state”, from *negarij*), *kolam* (“pond”), *kemeja* (“shirt”), *raket* (“racket”), *selimut* (“blanket”) and *sabun* (“soap”). Many of these words have been fully adapted to Indonesian phonology and morphology, sometimes acquiring new meanings in the process.

Dutch terms also contributed to administrative and legal vocabulary: *anggaran* (“budget”), *perkara* (“case”), *putusan* (“judgement”), *kontrak* (“contract”), *notaris* (“notary”), *faktur* (“invoice”), *organisasi* (“organisatie”), *partai* (“party”) and *universitas* (“university”). The prevalence of Dutch loanwords reflects the intensity and longevity of colonial rule and demonstrates how colonial languages can leave deep lexical imprints even after decolonization.

2.7.5 Other South Asian and Regional Influences

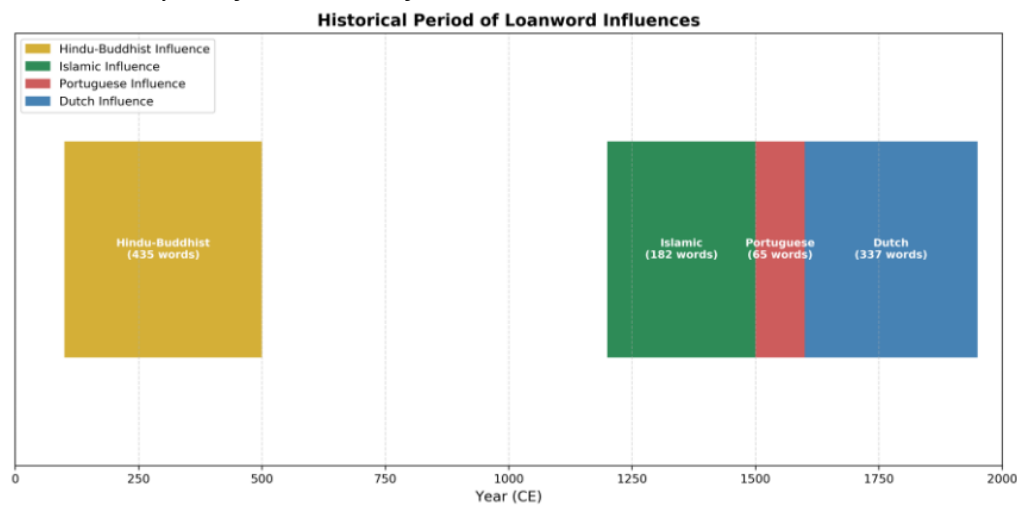
In addition to Sanskrit and Arabic, other South Asian languages have contributed vocabulary to Indonesian through historical trade and cultural exchange. Hindi-derived terms such as *ayah* (“caretaker”), *bibi* (“aunt”), and *garam* (“salt”) reflect these interactions (Ricklefs, 1993, p. 11). Although fewer in number, such borrowings demonstrate the broader network of linguistic exchange connecting Indonesia with South Asia.

Local languages within the Indonesian archipelago have also influenced the national lexicon. Languages such as Javanese, Sundanese, and Balinese have contributed numerous lexical items related to everyday life, cultural practices, and social organization. Examples include *sawah* (“rice field”), *gamelan* (a traditional musical ensemble), and *wayang* (shadow puppet theatre).

Furthermore, Chinese contact languages, particularly Hokkien, have contributed vocabulary related to trade, cuisine, and daily life, including words such as *mie* (“noodles”), *bakso* (“meatball”), and *teko* (“teapot”) (Makin, 2016, p. 5). These borrowings reflect the long history of commercial and cultural interaction between Chinese communities and Indonesian society.

These additional lexical influences illustrate the broader networks of cultural and linguistic exchange that contributed to the layered structure of the Indonesian lexicon analyzed in this study.

The historical layering of these linguistic influences is summarized in Figure 1, which provides a conceptual overview of the major phases of lexical borrowing that underpin the stratified structure of the Indonesian lexicon.

Figure 1.*Historical layers of loanword influence in Indonesian.*

The figure summarizes the major phases of linguistic influence associated with Hindu-Buddhist, Islamic, and European contact, which contribute to the stratified structure of the Indonesian lexicon.

3. Methodology

3.1 Data Collection

This study is based on a corpus of 1,402 lexical items in Indonesian whose etymological origins are associated with languages introduced through major religious and cultural contact. The dataset was compiled from authoritative sources, including the *Kamus Besar Bahasa Indonesia* (KBBI), historical dictionaries, and existing linguistic studies on loanwords in Indonesian.

Lexical items were selected based on their documented etymological origin and their relevance to domains historically associated with religious and cultural contact, including religion, governance, education, and everyday communication. While not all items are religious in meaning, their introduction into Indonesian is linked to languages transmitted through religious or colonial interaction.

3.2 Classification of Loanwords

The collected lexical items were classified according to their primary etymological source. The main categories include Sanskrit, Arabic, Dutch, Portuguese, Pali, Hindi and other South Asian languages, as well as regional and other sources.

In cases where etymology is contested or involves multiple possible origins, classification was based on the most widely accepted source in the literature. This approach ensures consistency while acknowledging the complexity of historical linguistic transmission.

To avoid misleading categorizations, language sources were not directly equated with specific religions. Instead, classifications reflect broader historical contexts of contact. For example, Dutch-derived loanwords are associated with colonial and institutional domains rather than exclusively with Christian influence.

The distribution of lexical items across these categories is presented in **Table 1**.

Table 1.

Distribution of Loanwords by Etymological Source in Indonesian

Source Language / Category	Number of Items	Percentage (%)
Sanskrit	438	31.2
Dutch (Colonial/Institutional)	340	24.2
Arabic	184	13.1
Regional and Other Sources	315	22.5
Portuguese	65	4.6
Hindi and Other South Asian	49	3.5
Pali	11	0.8
Total	1,402	100

The Regional and Other Sources category includes loans from regional languages, terms associated with indigenous belief systems and rare borrowings from Chinese and Persian sources. Some words, such as *sabun* (“soap”) and *kursi* (“chair”), have contested origins (Arabic vs. Dutch) and are assigned based on the balance of evidence. Although in the respect of *sabun*, for example, many etymological studies trace it further back to Greek through Arabic trade routes.

3.3 Analytical Approach

The analysis combines quantitative and qualitative methods. Quantitatively, the study examines the distribution of loanwords across etymological categories, identifying patterns of lexical stratification within the Indonesian lexicon. Percentages were calculated based on the total number of lexical items in the dataset.

Qualitatively, the study analyzes patterns of semantic adaptation, functional distribution, and sociolinguistic usage. Particular attention is given to how loanwords from different historical layers coexist and are used across religious, institutional, and everyday contexts.

This combined approach allows for a comprehensive examination of how successive waves of religious and cultural contact have contributed to the formation of a stratified lexical system in Indonesian.

The distribution shown in Table 1 provides the basis for the quantitative analysis of lexical stratification.

3.4 Data Availability and Replicability

The dataset used in this study was compiled from publicly available sources, including the *Kamus Besar Bahasa Indonesia* (KBBI) and published linguistic studies. While the full dataset is not included in this article, the criteria for data selection and classification are described in detail to ensure transparency and reproducibility.

Future research may build upon this dataset by applying similar classification criteria or expanding the corpus using additional lexical sources.

Representative examples of loanwords are provided in Appendix A.

4. Results and Discussion

4.1 Distribution of Loanwords by Etymological Source

The analysis reveals a clear stratification of loanwords in Indonesian, reflecting successive phases of religious and cultural contact. As shown in **Table 1**, Sanskrit-derived loanwords constitute the largest proportion of the dataset (31.2%), followed by Dutch (24.2%) and Arabic (13.1%), with smaller contributions from Portuguese, Hindi and other South Asian languages, and Pali.

This dominance suggests that early religious contact exerted long-term structural influence on the Indonesian lexicon, supporting cumulative models of lexical development rather than replacement-based models. The implication is that early religious contact continues to shape the core structure of contemporary Indonesian vocabulary, particularly in institutional and philosophical domains.

The substantial proportion of Dutch-derived loanwords reflects the enduring influence of colonial administration and institutional systems. This pattern indicates that later historical phases contribute functionally specialized lexical layers without displacing earlier ones. Theoretically, this supports models of layered lexical coexistence in language contact, where different historical influences map onto distinct functional domains. The implication is that colonial contact not only expanded the lexicon but also restructured domain-specific vocabulary in education, governance, and administration.

Arabic-derived loanwords, while representing a smaller proportion of the dataset, occupy a unique position due to their distribution across both religious and everyday contexts. This dual presence suggests that vocabulary introduced through religious transmission can extend beyond specialized domains into general communication. From a sociolinguistic perspective, this supports the view that religiously derived lexical items function as markers of identity and interaction. The implication is that religious contact contributes not only to lexical expansion but also to the formation of shared communicative repertoires across social contexts.

4.2 Patterns of Lexical Stratification

The findings demonstrate that lexical borrowing in Indonesian operates through cumulative layering rather than replacement. Each wave of contact—Hindu-Buddhist, Islamic, and European—has contributed new lexical resources without displacing earlier layers. As a result, the Indonesian lexicon exhibits a stratified structure in which multiple etymological layers coexist and interact.

This pattern suggests that language change in Indonesian does not follow a linear model of lexical substitution. Instead, it reflects an additive process in which successive layers accumulate and remain functionally differentiated. From a theoretical standpoint, this challenges replacement-based models of language change and supports stratification-based approaches in language contact theory. The implication is that lexical stratification constitutes a stable and enduring organizational principle of the lexicon rather than a transitional stage.

The differentiation of lexical layers across domains further reinforces this interpretation. Sanskrit-derived vocabulary is associated with formal and institutional registers, Arabic-derived terms with religious and interpersonal communication, and Dutch-derived forms with administrative and educational contexts. This distribution suggests that lexical stratification is not only historical but also functionally organized. The implication is

that speakers actively draw on different lexical layers to navigate communicative contexts, reinforcing the persistence of stratified lexical systems in contemporary usage.

4.3 Semantic and Functional Integration

In addition to their distribution across sources, loanwords in Indonesian exhibit patterns of semantic adaptation and functional integration. Many borrowed terms retain meanings close to their original forms, while others undergo semantic extension or specialization depending on context.

This variation suggests that lexical borrowing is not a uniform process but involves dynamic semantic adjustment within the recipient language. From a theoretical perspective, this supports models of semantic integration in language contact, where borrowed forms are adapted to fit local communicative needs. The implication is that the long-term survival of loanwords depends on their ability to align with existing semantic and functional structures within the language.

For example, Arabic-derived terms such as *iman* and *zakat* maintain strong religious associations, whereas other borrowings extend into broader domains of use. Similarly, Sanskrit-derived terms now permeate institutional vocabulary, often stripped of their original religious connotations. This pattern suggests that semantic flexibility facilitates the coexistence of multiple lexical layers. The implication is that semantic adaptation plays a central role in sustaining lexical stratification by enabling borrowed forms to remain relevant across changing contexts.

4.4 Sociolinguistic Implications

The coexistence of multiple lexical layers has important sociolinguistic implications. Loanwords in Indonesian function as resources for expressing register, identity, and social positioning. Speakers can draw on Sanskrit-derived formal vocabulary, Arabic-derived expressions of religiosity, and Dutch-derived institutional terminology to navigate different communicative contexts.

This pattern suggests that lexical stratification operates not only as a historical artifact but also as an active sociolinguistic resource. From a theoretical perspective, this aligns with sociolinguistic models that emphasize language as a repertoire of stylistic and social choices. The implication is that stratified lexical systems provide speakers with a flexible set of linguistic resources for constructing meaning and negotiating identity.

Furthermore, the ability to shift between lexical layers indicates that speakers actively engage with the historical depth of the language. This suggests that lexical stratification contributes to the richness and adaptability of Indonesian as a national language. The implication is that language contact not only shapes vocabulary but also enhances the expressive capacity of the linguistic system.

4.5 Theoretical Implications

The findings contribute to broader discussions in language contact theory and sociolinguistics by demonstrating how religious and cultural contact can produce durable lexical stratification within a national language. The Indonesian case shows that borrowing processes in multilingual societies often operate through accumulation rather than

replacement, resulting in layered lexical systems that reflect historical trajectories of interaction.

This suggests that existing models of language contact should place greater emphasis on long-term layering and functional differentiation rather than focusing primarily on borrowing as isolated events. Theoretically, this study supports an additive model of lexical development, in which successive layers remain active within the lexicon. The implication is that lexical stratification should be considered a central outcome of sustained language contact rather than a secondary effect.

By linking etymological origin with sociolinguistic function, this study also highlights the importance of integrating quantitative and qualitative approaches in the analysis of loanwords. Future research would benefit from moving beyond descriptive cataloguing toward analytical frameworks that account for both the distributional patterns of loanwords across domains and their functional role within the lexicon — an agenda for which corpus-based methodologies are particularly well-suited.

5. Conclusion

This study has examined the role of religious and cultural contact in shaping the Indonesian lexicon through a corpus-based analysis of 1,402 loanwords. The findings demonstrate that lexical borrowing in Indonesian reflects a stratified structure resulting from successive historical phases of contact, including Hindu-Buddhist, Islamic, and European influences.

The analysis shows that these layers do not replace one another but accumulate over time, forming a stable and functionally differentiated lexical system. Sanskrit-derived vocabulary constitutes a foundational layer associated with institutional and philosophical domains, while Arabic-derived terms extend across both religious and everyday contexts. Dutch and other European loanwords contribute an additional layer linked to administrative, educational, and modern institutional practices. Together, these patterns illustrate how different historical influences continue to coexist within contemporary Indonesian.

Beyond documenting the distribution of loanwords, this study advances research in language contact by demonstrating how religious interaction contributes to durable lexical stratification in a national language. The findings support a model of additive lexical development, in which successive layers of borrowing remain active and functionally differentiated rather than being replaced over time. This perspective highlights the importance of examining not only the origins of loanwords but also their interaction and use within a shared sociolinguistic system.

The study also underscores the value of corpus-based approaches for analyzing large-scale patterns of lexical borrowing. By combining quantitative distributional analysis with qualitative interpretation, it provides a more comprehensive account of how historical processes of contact shape contemporary language use.

A limitation of this study is its reliance on dictionary-based data and does not account for frequency variation in actual language use. Future research may address this limitation by incorporating frequency-based corpus analysis and examining variation across regions and speech communities. Further work could also explore how lexical stratification interacts with emerging forms of digital communication and language change.

In conclusion, the Indonesian lexicon illustrates how sustained religious and cultural contact can generate a layered and dynamic system of lexical resources. This case

contributes to broader discussions in sociolinguistics and language contact studies, demonstrating how historical processes continue to shape linguistic structure and social meaning in contemporary contexts.

Acknowledgements

The author sincerely thanks the instructors at the Indonesian Language Center for Cross-Cultural Communication (IMLAC) in Salatiga for their generous support and insightful feedback during this project. Gratitude is also extended to the archivists, librarians, and scholars of Indonesian linguistics, history, and religious studies, who provided essential resources and perspectives. All remaining errors are the authors own.

References

- Abdulkarim, A., Komalasari, K., Saripudin, D., Ratmaningsih, N., & Anggraini, D. N. (2020). Development of a unity in diversity-based Pancasila education textbook for Indonesian universities. *International Journal of Instruction*, 13(1), 371–384. <https://doi.org/10.29333/iji.2020.13125a>
- Alnizar, F., Mufidah, V. N., & Yani, Z. (2025). Framing faith and language: A critical discourse analysis of socio-religious etymologies in Indonesian lexicography. *LLT Journal: A Journal on Language and Language Teaching*, 28(1), 128–152. <https://doi.org/10.24071/llt.v28i1.9656>
- Alsaawi, A. (2020). The use of language and religion from a sociolinguistic perspective. *Journal of Asian Pacific Communication*, 32(2), 236–253. <https://doi.org/10.1075/japc.00039.als>
- Batais, S. S., & Wiltshire, C. R. (2022). Indonesian phonology and the evidence from loanword adaptation. In C. Shei & S. Li (Eds.), *The Routledge handbook of Asian linguistics* (pp. 245–260). Taylor & Francis. <https://doi.org/10.4324/9781003090205-17>
- Braithwaite, J., Braithwaite, V., Cookson, M., & Dunn, L. (2010). *Anomie and violence: Non-truth and reconciliation in Indonesian peacebuilding*. The Australian National University.
- De Casparis, J. G. (1997). *Sanskrit loan-words in Indonesian: An annotated check-list of words from Sanskrit in Indonesian and traditional Malay*. Badan Penyelenggara Seri NUSA, Universitas Katolik Indonesia Atma Jaya.
- Gil, D. (2024). Borrowing within Malayic: The role of exotericity. *Wacana: Journal of the Humanities of Indonesia*, 25(3), 480–530. <https://doi.org/10.17510/wacana.v25i3.1793>
- Gu, C. (2025). Searching for traces of Hindu/Buddhist heritage in the world's largest Muslim country: Indonesia's linguistic and semiotic landscape as a 'palimpsest'. *Religions*, 16(11), 1443. <https://doi.org/10.3390/rel16111443>
- Harahap, S. H., Perangin Angin, D., R., Br., Purba, J., I., S., Sembiring, S., H., Br., & Zaluku, D. (2025). Language and colonialism: A historical study on the development of Dutch loanwords in Indonesian. *Warisan: Journal of History and Cultural Heritage*, 6(1), 29–35. <https://doi.org/10.34007/warisan.v6i1.2665>
- Julul, A. A. H., Kwary, D., A., Rahmawati, N., M., & Sartini, Ni, W. (2020). Semantic adaptations of Arabic loanwords in the Indonesian language. *Mozaik Humaniora*, 19(2), 135–147. <https://doi.org/10.20473/mozaik.v19i2.14584>

- Lauder, A. F. (2010). Data for lexicography: The central role of the corpus. *Wacana: Journal of the Humanities of Indonesia*, 12(2), 219–242. <https://doi.org/10.17510/wjhi.v12i2.116>
- Mahfud, C., Astari, R., Kasdi, A., & Firdaus, W. (2021). Islamic cultural and Arabic linguistic influence on the languages of Nusantara: From lexical borrowing to localized Islamic lifestyles. *Wacana: Journal of the Humanities of Indonesia*, 22(1), 224–248. <https://doi.org/10.17510/wacana.v22i1.914>
- Makin, A. (2016). Unearthing Nusantara's concept of religious pluralism: Harmonization and syncretism in Hindu-Buddhist and Islamic classical texts. *Al-Jami'ah: Journal of Islamic Studies*, 54(1), 1–30. <https://doi.org/10.14421/ajis.2016.541.1-30>
- Manguin, P. (2021). Srivijaya: Trade and connectivity in the pre-modern Malay world. *Journal of Urban Archaeology*, 3, 87–100. <https://doi.org/10.1484/j.jua.5.123677>
- Musgrave, S. (2014). Language shift and language maintenance in Indonesia. In *language, education and nation-building* (pp. 87–105). Palgrave Macmillan. https://doi.org/10.1057/9781137455536_5
- Nesher, E. (2025). Religion as a context for language contact. *Journal of Religion and Linguistics*, 2(1), 38–50. <https://doi.org/10.31763/jorel.v2i1.15>
- Ricklefs, M. C. (1993). *A history of modern Indonesia since c.1300* (2nd Ed.). Red Globe Press. <https://doi.org/10.1007/978-1-349-22700-6>
- Rokib, M., Sodiq, S., & Mudzakkir, M. (2023). New Arabic loanwords in Indonesian dictionary. *Ijaz Arabi Journal of Arabic Learning*, 6(2), 526–535. <https://doi.org/10.18860/ijazarabi.v6i2.21090>
- Saputra, Y. W., Azmi, M., & Kurniawan, K. (2024). The use of andesite rocks in the Hindu-Buddhist kingdoms in Indonesia. *Yupa: Historical Studies Journal*, 8(2), 250–262. <https://doi.org/10.30872/yupa.v8i2.1514>
- Suri, N. S., Mascahaya, M., & Ritonga, F. H. (2020). Terms of address in Arabic in the Arab community in Medan city and their influence on Indonesian language development. *Arabi: Journal of Arabic Studies*, 5(1), 33–42. <https://doi.org/10.24865/ajas.v5i1.212>
- Wijaya, D. N., Wahyudi, D. Y., Umaroh, S. Z., Susanti, N., & Ertrisia, R. A. P. (2021). The toponymy of villages in Ambon Island: A historical and archaeological study. *Berkala Arkeologi*, 41(1), 89–108. <https://doi.org/10.30883/jba.v41i1.600>

Appendix A

Representative Loanwords by Etymological Source

Table A1Error! Reference source not found. provides representative examples of loanwords in Indonesian according to their etymological origin, along with their typical domains of use. The examples are drawn from the corpus analyzed in this study and are intended to illustrate patterns of lexical distribution and functional differentiation discussed in the main text.

Table A 1

Distribution of Lexical Origins by Region and Religious Prevalence

Lexical Origin	Associated Religion(s)	Primary Region(s)	Examples (from Corpus)	Notes on Use or Function
Sanskrit	Hinduism, Buddhism	Bali, Central Java	<i>Agama, Dharma, Upacara, Karma</i>	Institutional, ceremonial, and philosophical vocabulary.
Arabic	Islam	Java, Sumatra, Madura	<i>Iman, Doa, Masjid, Shalat, Zakat</i>	Deeply embedded in religious and daily vocabulary.
Dutch	Christianity (Protestantism), Colonial administration	Urban Java, Ambon	<i>Gereja, Sekolah, Katekismus, Pendeta</i>	Common in education, religious instruction, and governance.
Portuguese	Christianity (Catholicism)	Flores, Timor, Maluku	<i>Misa, Padre, Altar, Salib</i>	Predominant in Catholic-majority eastern regions.
Pali	<i>Theravāda</i> Buddhism	Limited (Buddhist minorities)	<i>Bhikkhu, Vinaya, Sangha</i>	Mostly found in textual or monastic contexts.