

Candidate Loanwords In Malaysian And Indonesian Online News: An Illustrative Comparison Of Forms And Semantic Domains

Alfi Rohmatul Asri^{1,*}, Rivalio Chandra Alfiansyah², Tiurmaida Sibarani³

^{1,2,3}Universitas Tidar, Magelang, Indonesia

Article Information

Article History:

Submitted: April 7, 2026

Revised: June 15, 2026

Accepted: June 29, 2026

Published: June 30, 2026

Keyword

student advisory information system;
Laravel; school data management; Waterfall
development; digital student support

Correspondence

E-mail:

alfi.rohmatul.asri@students.untidar.ac.id*

A B S T R A C K

Although Malay and Indonesian have longstanding histories of language contact, similarities among contemporary word forms alone cannot identify donor languages or pathways of transmission. This study offers an illustrative documentary comparison of 45 purposively selected word types: 15 associated with eight Malaysian news articles, 15 with five Indonesian news articles, and 15 attributed to Dutch in a separate secondary compilation published in 2023. The entries were examined using a simplified typology based on Haugen's borrowing categories and seven manually assigned semantic domains. All 45 preselected candidates were provisionally classified as loanwords; no loanblends or loanshifts were identified in these restricted lists. The largest domains were government/administration (14 entries) and social/community (11), followed by transport/trade and technology/media (six each), law/security (four), and household/technology and media/arts (two each). These figures represent counts of selected word types, not token frequencies or national proportions. The news articles were not sampled under comparable conditions, item-to-article extraction records were unavailable, and the proposed Dutch etymologies were not independently verified. The findings therefore provide a limited descriptive inventory and comparison of the three lists, rather than evidence of historical borrowing routes or broader differences between the languages. Replication requires traceable item-level sources, matched corpora, explicit coding procedures, and independent etymological verification.

Abstrak

Meskipun bahasa Melayu dan bahasa Indonesia memiliki sejarah kontak bahasa yang panjang, kemiripan bentuk kata saat ini saja tidak cukup untuk mengidentifikasi bahasa sumber atau jalur transmisi. Studi ini menyajikan perbandingan dokumenter terhadap 45 jenis kata yang dipilih secara sengaja: 15 kata dari delapan artikel berita Malaysia, 15 kata dari lima artikel berita Indonesia, dan 15 kata yang dikaitkan dengan bahasa Belanda berdasarkan kompilasi sekunder yang diterbitkan pada tahun 2023. Entri-entri tersebut dianalisis menggunakan tipologi sederhana yang didasarkan pada kategori peminjaman bahasa menurut Haugen serta tujuh ranah semantik yang ditentukan secara manual. Ke-45 kandidat yang dipilih tersebut untuk sementara diklasifikasikan sebagai kata serapan; tidak ditemukan adanya bentuk campuran (loan blends) maupun pergeseran makna akibat peminjaman (loan shifts) dalam daftar terbatas ini. Ranah dengan jumlah entri terbanyak adalah pemerintahan/administrasi (14 entri) dan sosial/kemasyarakatan (11 entri), diikuti oleh transportasi/perdagangan dan teknologi/media (masing-masing enam entri), hukum/keamanan (empat entri), serta rumah tangga/teknologi dan media/seni (masing-masing dua entri). Angka-angka ini menunjukkan jumlah jenis kata yang dipilih, bukan frekuensi kemunculan (token) ataupun proporsi secara nasional. Artikel-artikel berita tersebut tidak diambil dengan kondisi yang setara, catatan mengenai ekstraksi kata dari artikel tidak tersedia, dan etimologi bahasa Belanda yang diusulkan pun belum diverifikasi secara independen. Oleh karena itu, temuan ini hanya menyajikan inventaris deskriptif terbatas serta perbandingan dari ketiga daftar tersebut, bukan bukti mengenai jalur peminjaman historis atau perbedaan yang lebih luas antara kedua bahasa. Untuk melakukan replikasi, diperlukan sumber data yang dapat dilacak hingga ke tingkat kata, korpus yang sepadan, prosedur pengodean yang eksplisit, serta verifikasi etimologis secara independen.

This is an open access article under the CC-BY-SA license



1. Introduction

Lexical borrowing is a consequence of language contact, but similarities in written form cannot, on their own, establish a donor language or the direction of transfer. Haugen [1] distinguished loanwords, loanblends, and loanshifts as separate borrowing processes, while Tadmor [2] documented the varied linguistic sources of Indonesian vocabulary. These distinctions are particularly relevant to Malay and Indonesian, which share many lexical forms that may also resemble English or Dutch words. Synchronic comparison can identify orthographic correspondences and describe semantic domains; reconstructing borrowing pathways, however, requires independent historical or lexicographic evidence. The present study therefore separates the forms listed in its dataset from proposed source-language comparisons and avoids historical or causal conclusions that the available evidence cannot support.

A distinction between established lexical borrowing and occasional language mixing is important for interpreting isolated written examples. Poplack, Sankoff, and Miller [3] examined the social and linguistic conditions of lexical assimilation, while Poplack and Dion [4] showed why apparent stages of loanword development should not be presumed without longitudinal evidence. Cross-linguistic modelling by van Hout and Muysken [5], Tadmor et al. [6], and Zenner et al. [7] further indicates that borrowing patterns vary with lexical category, usage, and entrenchment. Thus, a collection of forms selected because they resemble foreign words cannot independently demonstrate how frequently borrowing occurs in a language.

The sociolinguistic meanings associated with borrowed vocabulary also require an evidential distinction. Zenner et al. [8] investigated the social meaning potential of loanwords using empirically grounded cases, whereas Schneider [9] discussed identity and contact in the development of new English varieties. Such research provides explanatory context rather than evidence about the intentions, identities, or linguistic awareness of the speakers who produced the news texts in the present sample.

Malay and Indonesian offer a valuable, albeit methodologically challenging, setting for comparison because they share a linguistic history but have experienced different patterns of institutional contact. Tan [10] investigated Malay-derived forms in Malaysian English newspaper texts, whereas Mohd Don et al. [11] examined English loanwords in a large corpus of Malay academic writing. Both the direction of borrowing and the linguistic register differ from those considered here. Bogunović [12] illustrates the value of systematically documented corpus-based lexical inventories, and Harahap et al. [13] situate Dutch-related Indonesian vocabulary in its historical context. None of these studies provides matched national samples against which the word lists examined here can be directly assessed. In particular, historical contact with Dutch administration does not, by itself, establish the etymology of a contemporary Indonesian word.

More closely related work on the Indonesian and Malay linguistic area reveals why simple English-versus-Dutch explanations are insufficient. Gil [14] analysed borrowing among Malayic varieties, Mahdi [15] traced older Austroasiatic-Austronesian contacts, and van Dam [16] critically reconsidered proposed Arabic borrowing routes into Indonesian. Studies based on early written materials by Hoogervorst [17] and Adelaar [18] show the type of historical documentation required to argue for particular pathways. Hardini and Grangé [19] likewise examined Indonesian lexical transfers involving French, including the possibility of intermediating languages. These findings do not validate any individual attribution in the present tables; instead, they demonstrate that borrowing histories may involve multiple layers of contact.

Related research also varies substantially in direction, medium, and analytical unit. Tan [20] examined Chinese-derived forms in Malaysian English, and P. K. W. Tan [21] considered Malay-origin items across English varieties; neither study directly measures English-to-Malay borrowing in news.

Research on Malaysian English based on social-media corpora [22] and a synthesis of social-media borrowing studies [23] further illustrates the influence of register and medium. Investigations of English-derived forms in Kelantan Malay [24] and Korean lexical items in Indonesian [25] use different kinds of linguistic evidence. Their findings inform research design without supplying comparable prevalence estimates for the selected lists examined here.

A central methodological concern is that lists of orthographically similar words may be interpreted as comparable evidence of borrowing even when they differ in source selection, unit of analysis, and etymological documentation. Studies of borrowing profiles underscore the importance of reliable historical identification [26], while a specialist reference on Indonesian and Malay loanwords offers a potential basis for future lexical verification [27]. The materials available for the present study, however, contain neither an item-level extraction record nor evidence of independent dictionary-based verification. Its contribution is accordingly more limited: it documents three differently sourced lists, reconciles their assigned borrowing classifications and semantic-domain counts, and specifies the conclusions that can reasonably be drawn from this evidence.

The gap addressed by this study is therefore not a previously unrecognised borrowing mechanism, but a mismatch between the level of evidence available in small lexical inventories and the scope of claims sometimes drawn from them. Biber [28] and Kilgarriff [29] establish the importance of corpus composition when interpreting distributional contrasts; Kilgarriff [30] also explains why linguistic data do not behave as random samples without a defensible sampling model. News discourse itself is shaped by topic and editorial news values, as analysed by Bednarek and Caple [31] and Bednarek [32]. A disciplined comparison must therefore treat differences in article selection and subject matter as potential explanations for the observed counts.

The analysis addresses four questions: (1) Which candidate borrowing-related forms appear in the selected Malaysian and Indonesian news-associated lists? (2) How are these forms provisionally classified using a simplified scheme based on Haugen? (3) How are the 45 selected word types distributed across seven assigned semantic domains? (4) Which interpretations of language contact remain plausible, and which cannot be verified from the documented sources? The comparison concerns selected word types only; it does not estimate borrowing rates across news media, establish historical transmission routes, test colonial causation, or assess speakers' awareness of lexical origins.

2. Method

This study used a qualitative, descriptive-comparative approach to analyse two sets of online-news examples alongside a supplementary published lexical compilation. The unit of analysis was the selected word type, defined as a distinct entry in one of the lists, rather than the number of occurrences of a word within an article. The study collected no participant responses, comprehension-test results, experimental observations, or complete-corpus frequency data. Borrowing categories were interpreted using the broad distinctions proposed by Haugen [1]. The analysis offers a descriptive comparison, not a matched-corpus assessment or a statistically representative account.

Data sources and selection. The Malaysian material comprised eight articles published between July 2024 and June 2026 by Free Malaysia Today [33], [34], [35], Utusan Malaysia [36], [37], and Astro Awani [38], [39], [40]. The Indonesian material comprised five articles from Antara News [41], CNN Indonesia [42], Kompas [43], Media Indonesia [44], and Republika [45], all cited as having been published in June 2026. Fifteen English-related candidate forms were selected from the Malaysian articles and 15 Indonesian forms with English cognates from the Indonesian articles. Although the article sources are identified, the available documentation contains no item-to-article extraction log, initial pool of candidate words, exclusion counts, or recorded explanation for selecting eight Malaysian articles and five Indonesian articles. The sample is therefore best described as purposive and illustrative rather than systematic, random, topic-matched, or independently reproducible.

Sampling and comparability rationale. Corpus-design research distinguishes a balanced, purposively constructed corpus from a convenience-based set of examples, emphasising coverage of relevant registers and situations [28]. Methods for comparing corpora also require explicit denominators and attention to different source compositions [29]. These principles serve here as criteria for evaluating what the available material can support; they were not implemented as a retrospective matched-sampling protocol. No unreported candidate searches, exclusions, or article-to-item mappings have been reconstructed.

Supplementary source and comparability. A third set comprised 15 Indonesian words attributed to Dutch in a secondary explanatory article published by Rizkita [46] in 2023. Unlike the first 30 candidates, these entries were not obtained through direct extraction from the sampled news articles. The three sets were presented as lists of 15 word types each, but equal list sizes do not make their sources or selection procedures comparable. The supplementary set provides contextual examples for discussing formal resemblances; it is neither a control group nor an independent sample of Indonesian news discourse. Its Dutch-language attributions remain claims reported in a secondary source, not independently verified etymologies.

Classification of borrowing types. The recorded classifications were interpreted using a simplified version of Haugen's framework [1]. A loanword retains recognisable formal material associated with a proposed source-language expression; a loanblend combines borrowed and recipient-language elements; and a loanshift involves semantic transfer or loan translation without an equivalent direct transfer of form. Because all three lists had been assembled from candidate borrowed forms, their composition introduced selection bias towards the loanword category. The documentation contains no systematic search for negative cases across the complete articles, formal procedure for resolving ambiguous classifications, independent second-coder assessment, or reliability coefficient. Thus, the classification of 45 entries as loanwords and none as loanblends or loanshifts applies only to the selected lists; it does not measure the distribution of borrowing mechanisms in the source publications.

Distinguishing written forms from phonological evidence. The tables record orthographic forms but no transcribed speech, acoustic observations, or speaker judgements. Accounts of phonological loanword adaptation have relied on explicit phonological models [47], perceptual experiments [48], and analyses of epenthetic vowels [49] and their phonetic realisation [50]. Those designs are not comparable to a manually selected list of written words; hence no phonetic mechanism or speaker-level adaptation process is inferred from the entries in this study.

Semantic domains and aggregation. The 45 entries were each assigned one of seven domains, as reported in Tables 1–3. Government/administration includes institutional, public-administration, educational-institution, and economic terms; law/security covers legal and enforcement-related vocabulary; social/community concerns interpersonal and collective life; transport/trade covers mobility, vehicles, and commercial logistics; technology/media concerns technical concepts and mediated communication; household/technology covers domestic objects and devices; and media/arts includes artistic and expressive vocabulary. These descriptions clarify the existing categories retrospectively; they do not constitute evidence that a codebook was established before analysis. Some words could plausibly belong to more than one domain, and no documented rule for resolving such ambiguity is available. Domain totals were calculated by counting the labels shown in the three tables, with each entry counted once in its assigned category.

Domain-definition considerations. Large-scale comparisons show that susceptibility to borrowing may differ across semantic fields [6]. Focused investigations of body-part terminology [51] and colour and flora/fauna vocabulary [52] illustrate how field boundaries and contact histories can influence interpretation. The seven domains used here are descriptive categories assigned to the available entries, not a validated cross-linguistic semantic taxonomy. The present totals should therefore not be

compared with the semantic-domain borrowing rates reported in studies using different coding schemes and complete lexical inventories.

Coding quality and reproducibility. Krippendorff [53] and Lombard et al. [54] distinguish explicit coding definitions from empirical evidence of intercoder agreement. O'Connor and Joffe [55] address the conditions under which qualitative researchers should report or interpret coding agreement, while Nowell et al. [56] describe auditability and transparency as components of qualitative trustworthiness. These studies inform the evaluation of the present documentation; they are not evidence that a second coder, reliability calculation, or prospective audit trail was actually used.

Ethical considerations and limits of interpretation. The sources were publicly available news reports and a published secondary explanatory article. No human participants were recruited, and the study makes no claim to have obtained formal ethics approval. Source articles are acknowledged bibliographically, and extended passages are not reproduced. Historical and psycholinguistic interpretations are necessarily restricted: formal resemblance does not establish a word's transmission history or what individual speakers know about its origin. Because publication periods, article topics, and sampling procedures were not matched, differences can be discussed only in relation to the three assembled lists.

Borrowing and code-switching boundaries. Malay-English bilingual data studied by Treffers-Daller et al. [57] demonstrate that the analysis of language alternation involves additional structural and interactional considerations. Because the current research did not collect spontaneous bilingual speech or speaker histories, it cannot decide whether every contemporary form reflects a particular stage of lexical integration, an individual alternation, or an historically established borrowing. Its borrowing labels remain provisional and specific to the recorded lists.

Data availability and auditability. Tables 1–3 present the 45 selected entries and their assigned domains, while Table 4 enables the reported counts to be checked. The available documentation does not include article-level concordance lines for the two news-associated sets, a candidate-screening log, annotated source texts, a documented coding manual, an independent coding audit, or item-by-item etymological checks. These records are acknowledged as unavailable rather than retrospectively reconstructed. The references identify the reported news articles where they remain accessible, but they do not establish which article supplied each individual entry.

3. Results and Discussion

3.1. Results

Table 1 presents 15 English-related Malay forms associated with the eight sampled Malaysian news articles. These selected examples do not constitute an exhaustive inventory of the articles' vocabulary. The available documentation provides neither article-specific source links nor contextual quotations for individual entries.

Table 1. Selected English-Related Loanwords in Malaysian Online News

No.	English Source	Malay Form	Domain	Category
1	record	Rekod	Law/Security	Loanword
2	syndicate	Sindiket	Law/Security	Loanword
3	activity	Aktiviti	Social/Community	Loanword
4	container	Kontena	Transport/Trade	Loanword
5	section	Seksyen	Government/ Administrati	Loanword
			on	
6	police	Polis	Law/Security	Loanword
7	case	Kes	Law/Security	Loanword
8	community	Komuniti	Social/Community	Loanword
9	crew	Kru	Transport/Trade	Loanword
10	boat	Bot	Transport/Trade	Loanword

No.	English Source	Malay Form	Domain	Category
11	party	Parti	Government/ Administrati on	Loanword
12	impact	Impak	Social/Community	Loanword
13	station	Stesen	Transport/Trade	Loanword
14	motorcycle	Motosikal	Transport/Trade	Loanword
15	campaign	Kempen	Government/ Administrati on	Loanword

Source: Eight Malaysian news articles [33], [34], [35], [36], [37], [38], [39], [40] (July 2024–June 2026). The categories and domains describe selected word types; the article-level provenance of individual entries was not documented.

Table 2 presents 15 Indonesian forms with English cognates alongside proposed Dutch orthographic comparanda. The Dutch forms are included as tentative points of comparison, not as evidence of confirmed historical borrowing routes. Independent etymological checks were not documented.

Table 2. Selected Indonesian Words with English Cognates and Tentative Dutch Comparanda

No.	English Cognate	Indonesian Form	Tentative Dutch Comparandum	Domain
1	community	komunitas	communititeit	Social/Community
2	discussion	Diskusi	discussie	Social/Community
3	criticism	Kritik	kritiek	Social/Community
4	pantomime	pantomim	pantomime	Media/ Arts
5	communication	komunikasi	communicatie	Technology/Media
6	commitment	komitmen	commitment	Government/ Administra tion
7	technology	teknologi	technologie	Technology/Media
8	transformation	transformasi	transformatie	Government/ Administra tion
9	operation	operasional	operationeel	Government/ Administra tion
10	verification	verifikasi	verificatie	Government/ Administra tion
11	projection	proyeksi	projectie	Technology/Media
12	economy	ekonomi	economie	Government/ Administra tion
13	industry	Industry	industrie	Technology/Media
14	mobility	mobilitas	mobilititeit	Transport/Trade
15	analysis	Analisis	analyse	Technology/Media

Source: Five Indonesian news articles [41], [42], [43], [44], [45] (June 2026). The Dutch forms are tentative comparanda rather than verified donor-language evidence. An item-to-article source mapping is unavailable.

Table 3 contains a supplementary set of 15 Indonesian words attributed to Dutch in a secondary explanatory compilation [46]. These words were neither extracted from the 13 sampled news articles nor independently checked against an etymological dictionary.

Table 3. Indonesian Words Attributed to Dutch in a Secondary Compilation (Supplementary Set)

No.	Attributed Dutch Form	Indonesian Form	Gloss	Domain
1	docent	Dosen	university lecturer	Government/ Administra tion
2	faculteit	Fakultas	faculty (of a university)	Government/ Administra tion
3	koelkast	Kulkas	refrigerator	Household/Technology
4	formatie	Formasi	formation/ structure	Government/ Administra tion
5	absentie	Absensi	absence/ attendance record	Government/ Administra tion

No.	Attributed Dutch Form	Indonesian Form	Gloss	Domain
6	emotie	Emosi	emotion	Social/Community
7	gage	Gaji	wage/salary	Government/Administration
8	identiek	Identic	identical	Social/Community
9	individu	Individu	individual	Social/Community
10	journaal	Jurnal	journal/log	Technology/Media
11	legendarisch	legendaris	legendary	Media/Arts
12	nummer	Nomor	number	Government/Administration
13	prestatie	Prestasi	achievement	Social/Community
14	ventilatie	Ventilasi	ventilation	Household/Technology
15	variatie	Variasi	variation	Social/Community

Source: Rizkita [46], detikJatim (2023), a secondary explanatory compilation. Attributions are retained as reported by that source and have not been independently verified.

Borrowing-type classification. All 45 preselected candidates were provisionally classified as loanwords under the simplified framework. No loanblends or loanshifts were identified within the three lists. Because candidate forms were chosen before classification, rather than through a comprehensive examination of the sampled articles' vocabulary, the result of 45 loanwords out of 45 entries is partly attributable to the selection procedure. It does not imply that loanblends or loanshifts are absent from the source articles or either language more broadly. Table 4 summarises the provisional borrowing classifications and reconciles the counts across seven assigned semantic domains.

Table 4. Provisional Borrowing Classes and Semantic-Domain Counts of 45 Selected Word Types

Category / Domain	Malay set	Indonesian cognate) set	(Eng.- Indonesian (Dutch) set	Total
Loanword	15	15	15	45 (100%)
Loanblend	0	0	0	0 (0%)
Loanshift	0	0	0	0 (0%)
Domain: Government/ Administration	3	5	6	14
Domain: Law/Security	4	0	0	4
Domain: Social/Community	3	3	5	11
Domain: Transport/Trade	5	1	0	6
Domain: Technology/Media	0	5	1	6
Domain: Household/Technology	0	0	2	2
Domain: Media/Arts	0	1	1	2

Note: The denominator is 45 selected word types (15 per list), not all word tokens or borrowing-related forms in the 13 news articles. The third list is derived from a secondary compilation; the totals are descriptive and not representative.

3.2. Discussion

The first two research questions concern the selected lexical forms and their provisional classification. All 45 candidates were assigned to the loanword category; none of the entries was classified as a loanblend or loanshift (Table 4). This uniform pattern does not demonstrate that loanwords predominate in contemporary news writing. The lists were assembled from presumed borrowing-related forms, and the complete lexical content of the sampled articles was not coded. Accordingly, the findings describe the classification of selected entries, not the relative occurrence of borrowing mechanisms across the 13 news articles. Similarity of form is directly observable, but identifying a donor language requires independent historical evidence.

The interpretation of the first result should distinguish a classification within a preselected list from an independent test of borrowing mechanisms. Empirical models of borrowability [5] and work on lexical entrenchment [7] require evidence about the population of possible forms, not just positive examples. In the present material, the absence of negative cases is a feature of the recorded selection process. It is therefore not possible to compare the observed 45-to-zero distribution with borrowing proportions established from broader lexical databases or systematic corpora.

The third research question addresses semantic domains. A recount of Tables 1–3 yields 14 government/administration entries, 11 social/community entries, six transport/trade entries, six technology/media entries, four law/security entries, two household/technology entries, and two media/arts entries. Transport/trade (five) and law/security (four) are comparatively prominent in the Malaysian list, whereas government/administration and technology/media each account for five entries in the Indonesian English-cognate list. These patterns may reflect the topics of the sampled articles and the breadth of categories assigned after selection, rather than stable differences in lexical usage. Diachronic corpus research on Korean borrowing by Oh and Son [58], for example, illustrates an analytical scale that distinguishes lexical types, tokens, and patterns of borrowing; such analyses cannot be reproduced from the present unmatched lists.

The domain counts likewise reflect both the lexical selections and the categories used to organise them. Comparative research has shown that semantic fields differ in their borrowing profiles [6], while a detailed study of body-part terms demonstrates how apparent resistance to borrowing can co-exist with social motivations for lexical transfer [51]. Gasser [52] examined borrowing in particular semantic fields using a different language-contact setting. The current government/administration and social/community totals cannot be interpreted as universal borrowing preferences, because neither a full vocabulary denominator nor independently validated domain boundaries are available.

The present study also differs in scope and evidential basis from Tan's [10] larger investigation of Malay-derived forms in Malaysian English newspaper texts. Since Tan examined borrowing in the opposite direction to the English-related Malay forms considered here, that study provides a useful methodological point of reference but not a directly comparable estimate of prevalence. Similarly, Bogunović [12] demonstrates the analytical value of a structured lexical database supported by explicit source documentation. Together, these approaches underscore the importance of traceable source records, which are unavailable for the 30 news-associated entries examined here.

Other Southeast Asian research provides a more specific methodological comparison. The corpus-based account of Korean-derived forms in Indonesian by Suhandano et al. [25] illustrates how a clearly identified medium supports the study of lexical distribution. Batais and Wiltshire [59] analysed Indonesian borrowing using formal phonological evidence, whereas the Kelantan Malay study by Nik Muhammad Naziman and Syed Jaafar [24] addressed spoken adaptation and perception. Tan [60] examined contact-related verb constructions in Malaysian English. These distinct research designs reinforce the need to separate orthographic description, frequency estimation, and claims about sound adaptation or morphosyntactic integration.

Mohd Don et al. [11] analysed English-derived forms, lexical distributions, and morphological integration in a substantial corpus of Malay academic-magazine writing. Although their findings inform the broader scholarship on borrowing, the difference in register and sampling design precludes direct numerical comparison with the 45 news-associated and supplementary entries considered here. Research on borrowing profiles likewise highlights the importance of etymological information and the interpretive difficulties posed by incomplete historical evidence [26]. The most meaningful comparison is therefore methodological, focusing on research design and evidential strength rather than estimated borrowing rates.

Differences in media are particularly consequential. Gulnazir and Salehuddin [22] traced lexical change using dated Malaysian English social-media material and concordance analysis, whereas

Zhang and Singh [23] synthesised studies of borrowing and lexical innovation on social platforms. These sources underscore the importance of identifiable time periods, usage contexts, and retrieval criteria. By contrast, the present news-associated lists do not contain complete article token counts or repeated observations of the same item. The appropriate conclusion is consequently limited to the forms assembled for descriptive comparison.

For the fourth question, Harahap et al. [13] discuss Dutch lexical influence, while Tadmor [2] identifies multiple sources of Indonesian borrowing. Neither verifies the Dutch comparanda in Table 2 or the attributions in Table 3. Although the reference edited by Jones [27] could aid future checks, no item-level verification was documented. Donor-language and transmission routes therefore remain unconfirmed.

Historical work further complicates the proposed donor-language contrast. Hoogervorst [61] relates Indian Ocean port vocabulary to multilingual contact networks; Hoogervorst [17] and Adelaar [18] use documentary wordlists to recover earlier forms, while Mahdi [15] documents contact across deeper linguistic periods. Van Dam [16] warns against treating repeatedly cited but weakly substantiated etymologies as established facts. Alongside the Indonesian study of French-origin vocabulary [19], this literature shows why a formal English cognate and a plausible Dutch intermediate form cannot, without additional evidence, establish the route followed by any particular entry in Table 2.

The lexical examples illustrate the distinction between formal observation and historical explanation. The Malay forms “sindiket” and “stesen” resemble the English words “syndicate” and “station”, while Indonesian “komunikasi” and “operasional” resemble the tentative Dutch comparanda “communicatie” and “operationeel”. These correspondences cannot establish when or through which language a form was transmitted, demonstrate regular phonological adaptation, or reveal its cognitive representation among speakers. Research on Indonesian-language borrowings into English has described different borrowing categories, including loanshifts, in another recipient-language context [62]; those findings cannot be assumed to apply to the present lists. No participant-based evidence of speakers' recognition or awareness of foreign lexical origins was collected.

Evidence for phonological integration and speaker awareness requires further differentiation. Paradis and LaCharité [47] developed an analysis of loanword adaptation, while Peperkamp et al. [48] experimentally investigated perceptual contributions to adapted forms. Studies of vowel insertion by Uffmann [49] and Kim and Kochetov [50] likewise rely on linguistic evidence absent from this dataset. Such work cannot be used to infer individual speakers' cognitive representations or establish sound-change rules merely from the spellings reproduced in Tables 1–3.

The study's principal limitations concern unequal sources, selection bias, and limited reproducibility. The eight Malaysian articles cover July 2024–June 2026, whereas the five Indonesian articles are dated June 2026; the third set was drawn from a separate secondary publication issued in 2023. Each set contains only 15 selected word types, without a denominator representing all eligible words or any record of token frequencies. The documentation also lacks inclusion and exclusion decisions, individual word-to-article contexts, rules for ambiguous semantic classifications, independent etymological checks, and intercoder agreement measures. These gaps preclude reliable national comparisons, confirmed historical attributions, or generalisations to the full news corpora. Moreover, the semantic domains overlap conceptually and have not been independently validated as mutually exclusive categories.

The absence of independent coding and item-level provenance also limits analytical dependability. Krippendorff [53] evaluates the properties of reliability measures, and Lombard et al. [54] emphasise transparent reporting of intercoder procedures. Qualitative-methods discussions by O'Connor and Joffe [55] and Nowell et al. [56] further clarify why an explicit decision trail matters even when inferential statistics are not the purpose of a study. These recommendations identify improvements for

future research; their inclusion does not establish that those practices were followed during the original analysis.

Future research could construct comparable Malaysian and Indonesian news corpora by aligning publication periods, topics, and article-selection criteria. An item-level extraction matrix and contextual concordance records would make lexical evidence traceable, while prospectively defined inclusion, exclusion, and negative-case rules would clarify sampling decisions. A documented coding manual should distinguish loanwords, loanblends, and loanshifts and specify how overlapping domains are resolved. Independent coding and systematic consultation of established etymological references could strengthen the assessment of coding consistency and proposed donor-language pathways. Complete article corpora would also enable separate analyses of word types and token frequencies rather than comparisons based on equally sized illustrative lists.

A stronger future design could triangulate news-derived records against historical materials and dictionaries, distinguishing contemporary attestations from donor-language histories. Studies within Malayic [14], historical Malay wordlist research [17], and accounts of borrowing through related language varieties [15] suggest how competing explanations can be tested rather than assumed. For present-day news, topic-matched sampling [31] and clearly defined corpus contrasts [30] would improve the interpretability of lexical-domain counts. These are prospective recommendations, not additional procedures or results claimed for this dataset.

4. Conclusion

This documentary study examined 45 deliberately selected candidate borrowing-related word types: 15 associated with Malaysian online news, 15 associated with Indonesian online news, and 15 Indonesian words attributed to Dutch in a separate secondary compilation. Under a simplified classification based on Haugen, all 45 entries were provisionally labelled loanwords; no loanblends or loanshifts were identified within the selected lists. Recounting the assigned semantic domains yielded 14 government/administration entries and 11 social/community entries, followed by six each in transport/trade and technology/media, four in law/security, and two each in household/technology and media/arts. The study contributes a documented account of these selected entries and distinguishes observable formal similarities from historical attributions that remain unverified. Because the datasets are not comparable samples, and item-level provenance and etymology were not independently confirmed, the results do not establish national borrowing rates, historical causation, or speakers' psychological awareness. More robust conclusions require matched sources, traceable item-to-article records, explicit coding procedures, and independent etymological verification.

References

- [1] E. Haugen, "The analysis of linguistic borrowing," *Language*, vol. 26, no. 2, pp. 210-231, 1950, doi: 10.2307/410058.
- [2] U. Tadmor, "Loanwords in Indonesian," in *Loanwords in the World's Languages: A Comparative Handbook*, M. Haspelmath and U. Tadmor, Eds. Berlin, Germany: Mouton de Gruyter, 2009, pp. 686-716.
- [3] S. Poplack, D. Sankoff, and C. Miller, "The social correlates and linguistic processes of lexical borrowing and assimilation," *Linguistics*, vol. 26, no. 1, pp. 47-104, 1988, doi: 10.1515/ling.1988.26.1.47.
- [4] S. Poplack and N. Dion, "Myths and facts about loanword development," *Language Variation and Change*, vol. 24, no. 3, pp. 279-315, 2012, doi: 10.1017/S095439451200018X.
- [5] R. van Hout and P. Muysken, "Modeling lexical borrowability," *Language Variation and Change*, vol. 6, no. 1, pp. 39-62, 1994, doi: 10.1017/S0954394500001575.
- [6] U. Tadmor, M. Haspelmath, and B. Taylor, "Borrowability and the notion of basic vocabulary," *Diachronica*, vol. 27, no. 2, pp. 226-246, 2010, doi: 10.1075/dia.27.2.04tad.
- [7] E. Zenner, D. Speelman, and D. Geeraerts, "Core vocabulary, borrowability and entrenchment: A usage-based onomasiological approach," *Diachronica*, vol. 31, no. 1, pp. 74-105, 2014, doi: 10.1075/dia.31.1.03zen.
- [8] E. Zenner, L. Rosseel, and A. S. Calude, "The social meaning potential of loanwords: Empirical explorations of lexical borrowing as expression of (social) identity," *Ampersand*, vol. 6, Art. no. 100055, 2019, doi: 10.1016/j.amper.2019.100055.

- [9] E. W. Schneider, "The dynamics of New Englishes: From identity construction to dialect birth," *Language*, vol. 79, no. 2, pp. 233–281, 2003, doi: 10.1353/lan.2003.0136.
- [10] S. I. Tan, "Lexical borrowing in Malaysian English: Influences of Malay," *Lexis*, no. 3, 2009, doi: 10.4000/lexis.629.
- [11] Z. Mohd Don, G. Knowles, and N. S. Mansor, "English loanwords in Malay: Identifying patterns in academic magazines," *GEMA Online Journal of Language Studies*, vol. 26, no. 2, pp. 283–298, 2026. [Online]. Available: <https://ejournal.ukm.my/gema/article/view/94248>
- [12] I. Bogunović, "A corpus-based approach to English loanwords: Introducing the database of English loanwords in Croatian," *Fluminensia*, vol. 35, no. 2, pp. 437–460, 2023, doi: 10.31820/f.35.2.1.
- [13] S. H. Harahap, D. R. Br Perangin-angin, J. I. S. Purba, S. H. Br Sembiring, and D. Zalukhu, "Language and colonialism: A historical study on the development of Dutch loanwords in Indonesian," *Warisan: Journal of History and Cultural Heritage*, vol. 6, no. 1, pp. 29–35, 2025, doi: 10.34007/warisan.v6i1.2665.
- [14] D. Gil, "Borrowing within Malayic: The role of exotericity," *Wacana: Journal of the Humanities of Indonesia*, vol. 25, no. 3, pp. 480–530, 2024, doi: 10.17510/wacana.v25i3.1793.
- [15] W. Mahdi, "Austroasiatic loanwords in Austronesian languages," *Wacana: Journal of the Humanities of Indonesia*, vol. 25, no. 3, 2024, doi: 10.17510/wacana.v25i3.1780.
- [16] N. van Dam, "Arabic loanwords in Indonesian revisited," *Bijdragen tot de Taal-, Land- en Volkenkunde*, vol. 166, no. 2–3, pp. 218–243, 2010. [Online]. Available: <https://www.jstor.org/stable/27868577>
- [17] T. G. Hoogervorst, "Seventeenth-century Malay wordlists and their potential for etymological scholarship," *Wacana: Journal of the Humanities of Indonesia*, vol. 25, no. 3, pp. 531–575, 2024, doi: 10.17510/wacana.v25i3.1782.
- [18] A. Adelaar, "Malay (and Javanese) loanwords in Frederick de Houtman's Malagasy wordlist," *Wacana: Journal of the Humanities of Indonesia*, vol. 25, no. 3, pp. 415–452, 2024, doi: 10.17510/wacana.v25i3.1781.
- [19] T. I. Hardini and P. Grangé, "An overview of Indonesian loanwords from French," *Indonesian Journal of Applied Linguistics*, vol. 6, no. 1, pp. 156–171, 2016, doi: 10.17509/ijal.v6i1.2749.
- [20] S. I. Tan, "Lexical borrowing from Chinese languages in Malaysian English," *World Englishes*, vol. 28, no. 4, pp. 451–484, 2009, doi: 10.1111/j.1467-971X.2009.01607.x.
- [21] P. K. W. Tan, "Malay loan words across different dialects of English," *English Today*, vol. 14, no. 4, pp. 44–50, 1998, doi: 10.1017/S026607840001052X.
- [22] N. Gulnazir and K. Salehuddin, "Uncovering the lexical variation and change in Malaysian English: A corpus-driven analysis," *GEMA Online Journal of Language Studies*, vol. 23, no. 4, 2023, doi: 10.17576/gema-2023-2304-02.
- [23] Z. Zhang and M. K. M. Singh, "The formation, usage, and linguistic features of Malaysian Standard English borrowed words in social media: A systematic literature review," *3L: The Southeast Asian Journal of English Language Studies*, vol. 32, no. 2, pp. 227–244, 2026, doi: 10.17576/3L-2026-3202-15.
- [24] S. Nik Muhammad Naziman and S. R. Syed Jaafar, "The adaptation strategies of English loanwords among Kelantan Malay dialect speakers," *3L: The Southeast Asian Journal of English Language Studies*, vol. 24, no. 4, pp. 128–142, 2018, doi: 10.17576/3L-2018-2404-10.
- [25] Suhandano, R. Febrina, A. Isti'anah, and H. W. Young, "Korean loanwords in Indonesian: A corpus-based study," *Wacana: Journal of the Humanities of Indonesia*, vol. 25, no. 3, pp. 453–479, 2024, doi: 10.17510/wacana.v25i3.1701.
- [26] M. de Heer, R. Blokland, M. Dunn, and O. Vesakoski, "Loanwords in basic vocabulary as an indicator of borrowing profiles," *Journal of Language Contact*, vol. 16, no. 1, pp. 54–103, 2024, doi: 10.1163/19552629-bja10057.
- [27] R. Jones, Ed., *Loan-Words in Indonesian and Malay*. Leiden, The Netherlands: KITLV Press, 2007.
- [28] D. Biber, "Representativeness in corpus design," *Literary and Linguistic Computing*, vol. 8, no. 4, pp. 243–257, 1993, doi: 10.1093/lhc/8.4.243.
- [29] A. Kilgarrieff, "Comparing corpora," *International Journal of Corpus Linguistics*, vol. 6, no. 1, pp. 97–133, 2001, doi: 10.1075/ijcl.6.1.05kil.
- [30] A. Kilgarrieff, "Language is never, ever, ever, random," *Corpus Linguistics and Linguistic Theory*, vol. 1, no. 2, pp. 263–276, 2005, doi: 10.1515/clt.2005.1.2.263.
- [31] M. Bednarek and H. Caple, "Why do news values matter? Towards a new methodological framework for analysing news discourse in Critical Discourse Analysis and beyond," *Discourse & Society*, vol. 25, no. 2, pp. 135–158, 2014, doi: 10.1177/0957926513516041.
- [32] M. Bednarek, "Voices and values in the news: News media talk, news values and attribution," *Discourse, Context & Media*, vol. 11, pp. 27–37, 2016, doi: 10.1016/j.dcm.2015.11.004.
- [33] Free Malaysia Today, "331 ditahan, pertaruhan haram Piala Dunia cecah RM2.62 juta," Jun. 29, 2026. [Online]. Available: <https://www.freemalaysiatoday.com/category/bahasa/tempatan/2026/06/29/331-ditahan-pertaruhan-haram-piala-dunia-cecah-rm2-62-juta>
- [34] Free Malaysia Today, "Kastam Johor rampas rokok seludup RM13.46 juta," Jun. 29, 2026. [Online]. Available: <https://www.freemalaysiatoday.com/category/bahasa/tempatan/2026/06/29/kastam-johor-rampas-rokok-seludup-rm13-46-juta>
- [35] Free Malaysia Today, "Ketua Pemuda PKR dakwa Anwar dinafi guna kemudahan kerajaan Johor," Jun.

- 29, 2026. [Online]. Available: <https://www.freemalaysiatoday.com/category/bahasa/tempatan/2026/06/29/ketua-pemuda-pkr-dakwa-anwar-dinafi-guna-kemudahan-kerajaan-johor>
- [36] Utusan Malaysia, "Bot tempatan guna kru warga asing ditahan," Jul. 2024. [Online]. Available: <https://www.utusan.com.my/nasional/2024/07/bot-tempatan-guna-kru-warga-asing-ditahan/>
- [37] Utusan Malaysia, "Jalan pulang kru MV Berkat Tuah masih buntu," Jun. 29, 2026. [Online]. Available: <https://www.utusan.com.my/berita/2026/06/jalan-pulang-kru-mv-berkat-tuah-masih-buntu/>
- [38] Astro Awani, "6 lelaki berpakaian hitam letupkan stesen minyak di Yala," Jun. 29, 2026. [Online]. Available: <https://www.astroawani.com/berita-dunia/6-lelaki-berpakaian-hitam-letupkan-stesen-minyak-di-yala>
- [39] Astro Awani, "PM Anwar lancar Malaysia Digital 2030, sasar 30 peratus sumbangan ekonomi digital kepada KDNK," Jun. 29, 2026. [Online]. Available: <https://www.astroawani.com/berita-malaysia/pm-anwar-lancar-malaysia-digital-2030-sasar-30-peratus-sumbangan-ekonomi-digital-kepada-kdnk>
- [40] Astro Awani, "PRN Johor: Ini satu lagi kenyataan yang bersifat angkuh, kata Kamil Munimz," Jun. 29, 2026. [Online]. Available: <https://www.astroawani.com/berita-politik/prn-johor-ini-satu-lagi-kenyataan-yang-bersifat-angkuh-kamil-munimz>
- [41] Antara News, "Pemanfaatan AI untuk bansos dan UMKM dorong peningkatan kesejahteraan," Jun. 17, 2026. [Online]. Available: <https://www.antaranews.com/berita/5612039/pemanfaatan-ai-untuk-bansos-dan-umkm-dorong-peningkatan-kesejahteraan>
- [42] CNN Indonesia, "PDIP mau lihat Jokowi naik kelas, bukan main level lokal di Lampung," Jun. 29, 2026. [Online]. Available: <https://www.cnnindonesia.com/nasional/20260629125536-32-1374613/pdip-mau-lihat-jokowi-naik-kelas-bukan-main-level-lokal-di-lampung>
- [43] Kompas, "Ketika warung jadul di Surabaya menjelma jadi ruang belajar," Jun. 29, 2026. [Online]. Available: <https://regional.kompas.com/read/2026/06/29/143158978/ketika-warung-jadul-di-surabaya-menjelma-jadi-ruang-belajar-yang>
- [44] Media Indonesia, "Tiga tahun beruntun berprestasi di kancah global, Holiday Inn & Suites Jakarta Gajah Mada raih Best Suite Hotel in Asia," Jun. 2026. [Online]. Available: <https://mediaindonesia.com/ekonomi/898957/tiga-tahun-beruntun-berprestasi-di-kancah-global-holiday-inn--suites-jakarta-gajah-mada-raih-best-suite-hotel-in-asia>
- [45] Republika, "Mobilitas lintas tengah Jawa meningkat, penumpang KA Kertanegara tembus 168 ribu," Jun. 2026. [Online]. Available: <https://rejogja.republika.co.id/berita/thdv7q349/mobilitas-lintas-tengah-jawa-meningkat-penumpang-ka-kertanegara-tembus-168-ribu>
- [46] N. Rizkita, "Kata serapan dari bahasa Belanda dalam bahasa Indonesia dan Jawa," detikJatim, Oct. 14, 2023. [Online]. Available: <https://www.detik.com/jatim/budaya/d-6981572/kata-serapan-dari-bahasa-belanda-dalam-bahasa-indonesia-dan-jawa>
- [47] C. Paradis and D. LaCharité, "Preservation and minimality in loanword adaptation," *Journal of Linguistics*, vol. 33, no. 2, pp. 379–430, 1997, doi: 10.1017/S0022226797006786.
- [48] S. Peperkamp, I. Vendelin, and K. Nakamura, "On the perceptual origin of loanword adaptations: Experimental evidence from Japanese," *Phonology*, vol. 25, no. 1, pp. 129–164, 2008, doi: 10.1017/S0952675708001425.
- [49] C. Uffmann, "Epenthetic vowel quality in loanwords: Empirical and formal issues," *Lingua*, vol. 116, no. 7, pp. 1079–1111, 2006, doi: 10.1016/j.lingua.2005.06.009.
- [50] K. Kim and A. Kochetov, "Phonology and phonetics of epenthetic vowels in loanwords: Experimental evidence from Korean," *Lingua*, vol. 121, no. 3, pp. 511–532, 2011, doi: 10.1016/j.lingua.2010.10.012.
- [51] K. Pattillo, "On the borrowability of body parts," *Journal of Language Contact*, vol. 14, no. 2, pp. 369–402, 2021, doi: 10.1163/19552629-14020005.
- [52] E. Gasser, "Borrowed color and flora/fauna terminology in Northwest New Guinea," *Journal of Language Contact*, vol. 12, no. 3, pp. 609–659, 2019, doi: 10.1163/19552629-01203003.
- [53] K. Krippendorff, "Reliability in content analysis: Some common misconceptions and recommendations," *Human Communication Research*, vol. 30, no. 3, pp. 411–433, 2004, doi: 10.1111/j.1468-2958.2004.tb00738.x.
- [54] M. Lombard, J. Snyder-Duch, and C. C. Bracken, "Content analysis in mass communication: Assessment and reporting of intercoder reliability," *Human Communication Research*, vol. 28, no. 4, pp. 587–604, 2002, doi: 10.1111/j.1468-2958.2002.tb00826.x.
- [55] C. O'Connor and H. Joffe, "Intercoder reliability in qualitative research: Debates and practical guidelines," *International Journal of Qualitative Methods*, vol. 19, 2020, doi: 10.1177/1609406919899220.
- [56] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules, "Thematic analysis: Striving to meet the trustworthiness criteria," *International Journal of Qualitative Methods*, vol. 16, no. 1, 2017, doi: 10.1177/1609406917733847.
- [57] J. Treffers-Daller, S. Majid, Y. N. Thai, and N. Flynn, "Explaining the diversity in Malay-English code-switching patterns: The contribution of typological similarity and bilingual optimization strategies," *Languages*, vol. 7, no. 4, Art. no. 299, 2022, doi: 10.3390/languages7040299.
- [58] Y. Oh and H. Son, "Lexical borrowing in Korean: A diachronic approach based on a corpus analysis,"

- Corpus Linguistics and Linguistic Theory, vol. 20, no. 2, pp. 407-431, 2024, doi: 10.1515/cllt-2022-0102.
- [59] S. Batais and C. Wiltshire, "Indonesian borrowing as evidence for Harmonic Grammar," *Journal of Linguistics*, vol. 54, no. 2, pp. 231-262, 2018, doi: 10.1017/S0022226717000317.
- [60] S. I. Tan, "Nativised prepositional verbs in Malaysian English from the perspective of language contact," *3L: The Southeast Asian Journal of English Language Studies*, vol. 19, no. 3, pp. 103-114, 2013.
- [61] T. G. Hoogervorst, "Sailors, tailors, cooks, and crooks: On loanwords and neglected lives in Indian Ocean ports," *Itinerario*, vol. 42, no. 3, pp. 516-548, 2018, doi: 10.1017/S0165115318000645.
- [62] A. Purwitasari, "Loanwords in English from local languages in Indonesia," *Ranah: Jurnal Kajian Bahasa*, vol. 14, no. 1, pp. 330-341, 2025, doi: 10.26499/rnh.v14i1.6655.