

## The Berawan Comparative Glossary: Motivation, Challenges and Undergirding Principles

**Jurgen M. Burkhardt**  
Universiti Malaysia Pahang  
juergburk@gmail.com

Received Date : 22 January  
2025

Accepted Date : 2 April 2025

Published Date : 27 June 2025

### ABSTRACT

*The paper describes the motivation for the glossary, the challenges in eliciting vocab on traditional lifestyle, flora and fauna, and the principles undergirding a unified orthography. Berawan is a language family comprised of four lects. The Berawan settlements are located along the Tinjar and Tutoh rivers, which are tributaries of the Baram River in the Malaysian state of Sarawak. The main output of this lexicography project was a comparative glossary with 1163 lexical items and for each of these items, all four lects are displayed side by side with English and Malay glosses. The Berawan lexemes are presented using a community approved unified orthography, which represents sounds that are pronounced the same way with the same graphemes across lects but also presents differences in a consistent way. The undergirding principles for the unified Berawan orthography are linguistic soundness, cross-lectal comparability, reproducibility, teachability, acceptability to all stakeholders and economical representation of sounds and words. It is hoped that the Comparative Berawan Glossary will become an impetus for the future production of Berawan dictionaries and literature.*

**Keywords:** Austronesian, Berawan, Borneo lexicography, glossary, orthography, diplomacy

Copyright: This is an open access article distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike 4.0 International (CC BY-NC-SA 4.0) license which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original work is properly cited.

### INTRODUCTION

The Berawan is a language family comprised of four lects, which are Batu Belah (BB), Long Teru (LTu), Long Jegan (LJg) and Long Terawan (LTn). Berawan is a member of the North Sarawak language family (Blust 1974). According to Eberhard, Simons and Fennig (2024), all the Berawan lects are considered endangered. The Berawan population as of 2020/2021 is 7508 speakers (Thomas Paran, personal communication, September 26, 2023). The Berawan settlements are found on the Tinjar and Tutoh rivers which are tributaries of the Baram River in the Malaysian state of Sarawak. On the Tutoh, the Batu Belah live in three villages and the Long Terawan in two

villages. On the Tinjar, the Long Teru have three villages and the Long Jegan have ten (Thomas Paran, personal communication, September 26, 2023).

The recently published Comparative Berawan Glossary / Glosari Komparatif Bahasa Berawan (Burkhardt, Burkhardt, & Ang, 2023), presents 1163 vocabulary items for each of the four Berawan lects side by side with English and Malay glosses. The lexical items are arranged according to semantic fields, for example ‘nature’, ‘flora and fauna’, ‘person or mankind’, ‘daily life’, etc. Each semantic field is further divided into subfields. The field ‘flora and fauna’, for example, is subdivided into the subfields ‘land animals’, ‘birds and bats’, ‘fish and other river creatures’, ‘forest and plants’, and ‘fruits’. The semantic fields and subfields are drawn from Moe (2001). Table 1 is an excerpt from the comparative glossary:

**Table 1** – Excerpt from Burkhardt, Burkhardt, & Ang (2023, p. 39)

Batu Belah	Long Teru	Long Jegan	Long Terawan	Malay	English	Item
<b>Flora and fauna (Flora dan fauna)</b>						
<b><i>Birds and bats (Burung dan kelawar)</i></b>						
manok	manok	manâwk	manô'	burung	bird	159#
pelakkéh	pelakkéh	pelakkây	pelakkêh	helang	eagle / hawk	165#
temenngang	tebenngang	temenngang	belenngang	burung enggang	Rhinoceros hornbill	166#
jelinât	jelinât	jelênât	luka'	burung enggang hitam	Malayan black hornbill	167#
tekukkoh	tekukkoh	sekokkâw	kuku	burung enggang jambul putih	White-crowned hornbill	168#
mattoy	mattoy	mattoy	mattoy	burung enggang papan	Great Hornbill	169#
tepiyon	tepiyon	tepiyon	tekiyon	burung enggnag kelingking	Oriental Piped Hornbill	170#

While the comparative glossary is the heart of the so named publication, the book also contains a brief sketch of the historical development of the Berawan lects, a summary description of the consonant and vowel systems of the Berawan lects, and a spelling guide with examples of Berawan words across lects. The sketches and descriptions are all in English and Malay, due to the bilingual character of the book. Likewise, the glosses for the Berawan lexemes and the indices are presented in English and Malay.

The following sections describe the motivation for producing the comparative glossary, the challenges in the elicitation process of Berawan vocabulary, and the principles that undergird the writing of Berawan words.

The following notational conventions are used in this article: For the representation of Berawan words and sounds, phonetic transcriptions are presented in square brackets ([...]), phonemes are set in forward slashes (/.../) and orthographic notations appear in italics, for example BB/LTu *manok* /manok/ [manək] ‘bird’. Burkhardt (2014) uses the breve for the phonemic notation of short vowels (e.g. /ă/). However, in this article, they are presented with a circumflex instead (e.g. /â/). Acronyms used to refer to proto-languages are as follows: PAn stands for Proto-Austronesian, PMP for Proto-Malay-Polynesian, and PWMP for Proto-West-Malayo-Polynesian. All PAn, PMP and PWMP items are taken from Blust, Trussel, & Smith (2023).

## METHODOLOGY

### *Motivation for the Glossary*

The groundwork for the glossary is Burkhardt’s (2014) dissertation entitled ‘The reconstruction of the phonology of Proto-Berawan’. The main goal of the dissertation was to reconstruct the system of phonemes of Proto-Berawan, the assumed ancestor language of the currently spoken Berawan lects. For that purpose, the sound systems of the individual Berawan lects needed to be analysed and described. Furthermore, sufficient vocabulary from all Berawan lects had to be collected to be able to reconstruct a pool of proto lexemes that is big enough to draw valid conclusions about phonological innovations that are shared by all Berawan varieties but do not occur in other languages of the North Sarawak language family. For example, all Berawan lects, but no other North Sarawak language, exhibit a change from PAn/PMP/PWMP \*-b to -m, as in PAn \*qaseb > all Berawan *cam* ‘smoke’, PMP \*e(ŋ)keb ‘cover’ > all Berawan *kâm* ‘cover’, and PWMP \*eleb > all Berawan *lâm* ‘knee’ (Burkhardt, 2014, p. 304-305).

One result of the dissertation was the reconstruction of 490 Proto-Berawan lexemes. The dissertation had since been made available online on Research Gate and this was also made known to Berawan speakers so that they could access the outcome of this work. On the other hand, the dissertation is of limited benefits to the Berawan since it had to be written in a very technical (linguistic) jargon, and the corpus of reconstructed items and the Berawan language items on which they are based is rather limited, even though the number of lexical items used and reconstructed is 2.5 higher than the typical 200-items Swadesh list used for such endeavours. Nevertheless, the items used across Berawan lects for these reconstructions are a good basis for the development of a glossary and a future dictionary.

Another result of the dissertation was the identification of the phoneme system for each Berawan lect. This forms the basis for the development of Berawan orthographies that allow for an accurate written representation of Berawan lects, usable for Berawan speakers to read and write their respective lect.

Due to the comparative nature of the dissertation, sound correspondences were established across Berawan lects, as shown in Table 2.

**Table 2:** Sound correspondences across Berawan lects (Burkhardt, 2014, p. 385)

<b>Proto-Berawan</b>	<b>BB</b>	<b>LTu</b>	<b>LJg</b>	<b>LTn</b>	<b>Gloss</b>
*accih	accih	accih	cey	accîh	one
*akkuh	akkuh	akkuh	akkew	akkûh	ashes
*appok	appok	appok	appâwk	appô'	dust

Therefore, a comparative glossary was envisioned because it allows to list such items in a clear and comparable manner and thus aids speakers of Berawan lects to become aware of and learn about the similarities and differences with other lects. This has the potential of raising cross-dialectal intelligibility among Berawan speakers. According to Thomas Param (pers. comm.), cross-lectal intelligibility was higher among the Berawan in the past. In present times, it has decreased mainly due to two factors. On the one hand, many Berawans have moved out from their traditional settlements to towns and trade-centres in search for employment and many of them entered ethnically mixed marriages (Burkhardt & Burkhardt, 2017, p. 58). Their children are usually raised with Malay or English as the language of the home and the Berawan lect of one of their parents has become their heritage language, of which they acquire a basic knowledge at best. Second, there are less intermarriages across Berawan groups nowadays compared to the past. This has decreased inter-dialectal intelligibility. To reach true comparability, the lexical items presented in the glossary need to be presented in a unified orthography that represents sounds that are the same with the same orthographic symbols consistently. Thus, a motivation of the comparative glossary, on the receptive side, is to help raise intelligibility across Berawan lects.

On the productive side, a motivation for the compilation of the glossary is to provide an impetus for Berawan speakers to write their lects. For this purpose, the glossary does not only list lexical items in a consistent united orthography but contains a spelling guide where every Berawan vowel and consonant sound is listed with examples across lects.

Another perhaps more urgent motivation for the comparative glossary, due to the status of Berawan language endangerment, is its focus on the documentation of vocabulary that relates to the traditional riverine lifestyle of the Berawan and the flora and fauna that surrounds their traditional dwellings. Therefore, the collection of lexical items was increased from the dissertation's 490 items to 1163 items per lect.

On the other hand, the comparative glossary is not only geared towards Berawan speakers and readers, but also towards the international linguistic community. For this purpose, a comparative summary of the Berawan sound and phoneme system is explained in a separate section, as well as the methodology employed for the elicitation of the glossary items and the development of the unified orthography. Furthermore, Berawan items listed may provide helpful evidence for historical linguists to reconstruct new Proto-West-Malayo-Polynesian proto-lexemes.

To capture a regional audience (Malaysia, Indonesia, Nusantara) as well as an international one, the introductory section of the glossary, the glosses for the Berawan lexemes and an alphabetical index of glosses is provided both in Malay and English.

### ***Challenges in Elicitation of Berawan Vocabulary***

The language consultants for the elicitation of traditional Berawan vocabulary were Berawan native speakers 50 years of age or older. The elicitation of the vocabulary of daily activities did not prove too difficult. On the other hand, a semantic field where it was difficult for language consultants to recall specific lexemes was fauna. For example, when types of hornbills were elicited, the initial results obtained were different in quantity across lects. In two lects, the Berawan names for the rhinoceros hornbill and the oriental piped hornbill were obtained. In a third lect, the answer also included the name for the Malayan black hornbill and in a fourth lect, the name for the great hornbill could be elicited besides the names for the other three hornbill species. In a second elicitation round, the author obtained the names for the Malayan black and the great hornbill from the first two lects by prompting them with the respective item from the other two lects. In addition, pictures of each type of hornbill were shown to the informants to verify the accuracy of the respective Berawan term per lect.

The historical sound changes described in Burkhardt (2014) usually helped to predict the correct term across lects. For example, LJg *jelénat* ‘Malayan black hornbill’ would be expected to be BB and LTu *jelinat*, since they retain Proto-Berawan \*i in the penultimate syllable, whereas in LJg, the penultimate high front vowel gets lowered to the mid-front vowel *é* across an intervening consonant (*n* in this case), if the vowel in the ultima is non-high (*a* in this case), a phenomenon termed regressive vowel-height harmony (Burkhardt, 2014, p. 81). The prediction proved correct and the BB and LTu language consultants responded with *jelinat*.

Another semantic field where the recall of indigenous vocabulary proved challenging were traditional ornaments such as traditional earrings, necklaces, bracelets and ornamental hats. Prompting from one dialect to another was also used in this case to elicit the respective terms across lects. In one instance, the vernacular word for traditional long female earrings could not be recalled anymore, so the respective spot in the glossary was left empty. As less and less people wear traditional ornaments, the vernacular vocabulary for these items slowly fades away. Similarly, it is also not surprising that loss of biodiversity and the fading away of flora and fauna vocabulary go hand in hand.

As it came to indigenous words for four different types of gongs, the lexical items were readily remembered, which are (1) *sanang*, (2) *gennéng*, (3) *tawâk* in all Berawan lects, and (4) BB/LTu *gung*, LJg *agewng*, LTn *agung*. However, when pictures of the four types of gongs were shown to the language consultants, there was no general agreement among them which lexeme

refers to which gong. Therefore, these gongs are merely listed with the general meaning ‘gong’ in the respective Malay and English gloss columns in the comparative glossary.

## DISCUSSION

### *Principles undergirding the writing of Berawan words*

#### *Linguistic soundness*

To make the lexical items across Berawan lects truly comparable, a unified orthography was adopted. It is based on the identification of the vowel and consonant phonemes of each Berawan lect, an outcome of Burkhardt (2014). This determines the number of phonemes the Berawan lects possess in sum and how many orthographic symbols are needed to represent these phonemes. Ideally, an orthography should be “linguistically sound, acceptable to all stakeholders, teachable and easy to reproduce” (Cahill & Karan, 2008, p. 3). A linguistically sound orthography ideally exhibits a one-to-one correspondence of phonemes and graphemes, whereby a grapheme can be a monograph or polygraph, the latter typically being a digraph (e.g. *ng* or *iw*) or trigraph (e.g. *nng* or *iew*). This avoids underrepresentation on the one hand and overrepresentation on the other. Underrepresentation occurs when one orthographic symbol represents two or more phonemes. This is the case for the letter *e* in the current Malay orthography. On the one hand, it represents phoneme /é/ [ɛ], referred to as ‘e taling’, as in *perang* /péran/ [peran] ‘brown’. On the other hand, it represents schwa /e/ [ə], referred to as ‘e pepet’, as in *perang* /peran/ [pəran] ‘war’. An example for overrepresentation are the German orthographic symbols *f* and *v*, which both represent phoneme /f/, as in *voll* /fol/ ‘full’ and *faul* /faul/ ‘lazy’. For vital languages with millions of speakers, like Malay and German, this does not pose a problem. For endangered languages, however, under- or overrepresentation should be avoided as far as possible, since it creates inaccuracies in written language documentation and it potentially complicates language acquisition by young Berawans who wish to learn their heritage language, but whose mother tongue is Malay or English. Therefore, the unified Berawan orthography uses *é* and *e* as separate graphemes for phonemes /é/ [ɛ] versus /e/ [ə], for example in BB/LTu/LJg *cén* /cɛ́n/ [cɛ́n] ‘strength’ versus *cen* /cɛ́n/ [cɛ́n] ‘animal’ (Burkhardt, Burkhardt, & Ang 2020, p. 10).

#### *Cross-lectal Comparability*

To be able to compare the Berawan lects, there needs to be not only a one-to-one correspondence between grapheme and phoneme within a lect, but also, as far as possible, across lects. Therefore, long monophthongs are orthographically represented with a single vowel symbol without diacritics, which are *a*, *i*, *o*, *u* across all Berawan lects. The exception is *é* /é/ [ɛ], which needs a diacritic to distinguish it from schwa (*e* /e/ [ə]). Short monophthongs are represented consistently with a diacritic across lects, which are *â*, *ê*, *î*, *ô*, *û*. The exception is schwa *e*, which does not need a diacritic since it is always short (Schwa is far more frequent than *é*). Likewise, diphthongs are written the same way across lects. Long diphthongs are written *ay* and *aw*, as *ay* in all Berawan *paray* ‘rice plant’ and as *aw* in BB/LTu/LTn *mutaw* and LJg *motaw* ‘tired’. Short diphthongs are written *ây* and *âw*, as *ây* in LTn *lây* ‘upper arm’ and LJg *parâyk* ‘to shave’, and as *âw* in LTn *gimâw* ‘root’ and LJg *tâwng* ‘top end (of a tree)’ (Burkhardt, Burkhardt, & Ang 2020, p. 10-11), with the circumflex on the base vowel like the diacritic used for short monophthongs. There is one exception to the one phoneme – one grapheme approach, which is Long Jegan word-final *-ng*. The



grapheme is used in this position to refer to [ŋ] and [ɲ], reflecting the intuition of Long Jegan speakers. However, this does not pose a problem, since the two sounds occur in complementary distribution in word-final position: [ɲ] occurs after diphthongs with a high front offglide [ɪ], as in *akeyng* /akeyɲ/ [kakə'ɲ] ‘left (side)’ and *kérâyng* /kérâyɲ/ [kəra'ɲ] ‘new’, but in other environments, only [ŋ] occurs, as in *tolang* /tolaŋ/ [tolaŋ] ‘bone’, *ameng* /aməŋ/ [aməŋ] ‘dump’, and *monâwng* /monâwŋ/ [mona'wŋ] ‘mouth’ (Burkhardt, 2014, p. 185). Table 3 shows how cross-lectal comparability is achieved by providing one-to-one correspondences between vowel phonemes and graphemes across lects. The graphemes are represented the same way as the phonemes.

Table 3 Cross-lectal comparability of Berawan vowel phonemes / graphemes

	<b>Monophthongs of default length (mid-long to long)</b>	<b>Diphthongs of default length (mid-long to long)</b>	<b>Extra-short vowels</b>	<b>Extra-short diphthongs</b>
<b>BB/LTu</b>	a é i u o	ay, oy, uy, aw, iw	â, e	(ew)
<b>LJg</b>	a é i u o	ay, oy, uy, aw, iw	â, e	ây, âw, ey, ew
<b>LTn</b>	a é i u o	ay, oy, uy, aw, iw	â, e, ê, î, ô, û	ây, âw, ey, ew

z

### ***Reproducibility and Teachability***

All graphemes should be easily reproducible. Nowadays, computers and smartphones are widely used in Malaysia and the Berawans are no exception to this. Therefore, graphemes should be used that are reproducible on such devices. The graphemic symbols used in the unified Berawan orthography are easy to type on computers and smartphones with an English or Malay language setting. The diacritics that were chosen to represent short vowels are found in the list of vowels with diacritics that appears if a key is held down on a virtual smartphone keyboard for about a second. The smartphone list that appears for *a*, for example, contains à, á, â, ã, æ, ã, â, and ã. In academic transcriptions of classical manuscripts, the breve (*ă*) is frequently used to mark short vowels (Zui, June 24, 2023). However, that symbol is not included in the above-mentioned list. Therefore, the circumflex (*â*) was chosen to represent the low short vowel since it looks like an upside-down version of the breve.

To be teachable, graphemes that are also used in Malay, the national language, were chosen for the same sounds: *ng* for /ŋ/ [ŋ], *ny* for /ɲ/ [ɲ], *c* for /cç/ [cç] and *j* for /[j]/ [j], and *e* for Schwa /ə/ [ə] etc. For the mid front vowel /é/ [ɛ], *é* was selected, which is the same symbol that was used for that sound in the Malay orthography before its 1972 reform (Ismael, 2000). Furthermore, the one-to-one correspondence between Berawan phonemes and graphemes facilitates teachability.

### *Acceptability to all Stakeholders*

In the development of a national or regional language for usage in government, courts, business and education, there are many stakeholders including politicians, lawyers, educators and not to forget ordinary speakers of the language. Since Berawan lects do not have such a status, the stakeholders are, besides ordinary Berawan speakers, the Berawan leaders on a regional and village level (Penghulus and village chiefs), other influential Berawan speakers such as educationists and clergy, and the Berawan Sarawak association (Persatuan Berawan Sarawak). To ensure acceptability of the unified orthography in the Berawan community, an orthography workshop was held in Miri in June 2018, organized by the Berawan community. About 35 Berawans were present, among them current and former village chiefs, and members of the Berawan cultural association. The author was invited as a linguistic consultant. The unified Berawan orthography reflects decisions that were made in that workshop. So, for example, the workshop participants decided to represent the glottal stop with an apostrophe ('), not with letter *q*, as in BB/LTu/LTn *tinyo'* and LJg *tényo'* 'banana', not *tinyoq* / *tényoq*.

### *Economical representation*

The Merriam-Webster dictionary (n.d.) defines the term economical as “operating with little waste or at a saving”. An economical orthography avoids strings of graphemes that are longer than necessary. For this purpose, geminate alveopalatal and velar nasals /ɲ:/ and /ŋ:/, are represented as *nny* and *nng* respectively, instead of *nyny* and *ngng*. Thus, one letter is saved in each instance. Another issue is the representation of short versus long vowels. Instead of writing them *â* (short) versus *a* (long), they could be written *a* (short) versus *aa* (long). The advantage of the latter approach is that this would not require any diacritics, except for the mid-front vowel (é). The disadvantage is that one more letter is needed to represent long vowels. In Long Terawan, there is a phonemic length contrast for all vowels except schwa, so if the second alternative is employed, it would be *a* - *aa*, *é* - *éé*, *i* - *ii*, *o* - *oo*, and *u* - *uu*. In the other three lects, however, there is a length contrast only for low central vowels (*a* - *aa*), but not for any other vowel. So, in a unified orthography, the non-central vowels, which are long by default in the ultimate syllable, would have to be written with a doubling of the grapheme (*éé*, *ii*, *oo*, *uu*) even in the absence of phonemic contrast. This would not be economical.

Frequency of occurrence is another important criterion to arrive at an economical orthography. It is more economical to use diacritics for infrequent sounds than for frequent ones. The unified Berawan orthography accomplishes that by marking short vowels, which occur infrequently, with a diacritic, but not the far more frequent long vowels. The exception to this is the mid-front vowel, that needs a diacritic for the long vowel (é), and another one for the short one (ê). Furthermore, frequently occurring schwa (*e*) does not need a diacritic because it is always short. The historical analysis of Berawan also lends credence to this choice. The long Berawan vowels (or rather vowels of default length) are usually the ones that retain the proto form, whereas the short vowels (or rather extra-short vowels) are innovations that are only found in a limited set of lexical items. Furthermore, the extra-short vowels marked in bold in table 4 are only found in the Long Terawan lect. Table 4 contrasts uneconomical grapheme choices with the economical ones adopted in the Comparative Berawan Glossary.



**Table 4** Uneconomical vs. economical representation of Berawan graphemes

	Phoneme	Grapheme (uneconomical)	Grapheme (economical)
<b>Consonants</b>	ñ:	nyny	nny
	ŋ:	ngng	nng
<b>(Mid-)long vowels</b>	a:, é:, i:, o:, u:	aa, éé, ii, oo, uu	a, é, i, o, u
<b>Extra-short vowels</b>	a, e, é, i, o, u	a, e, é, i, o, u	â, e, ê, î, ô, û

## CONCLUSION

The first motivation for the Comparative Berawan Glossary is to document comparable vocabulary from all Berawan lects with an emphasis on endangered vocabulary in a unified orthography, supplemented by a spelling guide and description of the Berawan sound system to aid the Berawans in maintaining their lects, in raising cross-lectal intelligibility and in motivating them to write and further document their lects. The second motivation is aimed at linguists to provide them with a model for developing unified orthographies for other language families and with a useful set of language data for the reconstruction of PWMP forms. The comparative glossary is a balancing act between making it comprehensible and applicable for Berawan speakers and at the same time insightful for linguists. Challenges in elicitation were encountered when it came to ornamental items associated with traditional lifestyle and local fauna. A second round of elicitation with cross-lectal prompting through items collected during the first elicitation session generally proved successful to collect the missing lexical items across lects. The Berawan lexemes are presented in a unified orthography, which renders sounds that are pronounced the same way with the same graphemes across lects but also presents differences in a consistent way. The undergirding principles for the unified Berawan orthography are linguistic soundness (particularly a one-to-one correspondence between phoneme and grapheme), cross-lectal comparability, reproducibility and teachability, acceptability to all stakeholders and economical representation of sounds and words. It is hoped that the Comparative Berawan Glossary will serve as a stimulus to produce full-fledged Berawan dictionaries and Berawan literature in the future.

## REFERENCES

- Blust, R. 1974. *The Proto-North Sarawak vowel deletion hypothesis*. [Unpublished doctoral dissertation]. University of Hawai'i at Manoa.
- Blust, R., Trussel, S., & Smith, A. D. 2023. *CLDF dataset derived from Blust's "Austronesian Comparative Dictionary"* (v1.2) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7741197>
- Burkhardt, J.M. 2014. *The reconstruction of the phonology of Proto-Berawan*. [Doctoral dissertation, Goethe University]. DOI:10.13140/RG.2.2.17517.08160
- Burkhardt, J.M., Burkhardt, J.L., & Ang, L.H. 2020. One Orthography, Four Lects: The Unified Berawan Orthography. *Journal of Borneo-Kalimantan*, 6(1), pp. 1-21. <https://doi.org/10.33736/jbk.2411.2020>
- Burkhardt, J.M., Burkhardt, J.L., & Ang, L.H. 2023. *Comparative Berawan glossary – Glosari komparatif Bahasa Berawan*. Universiti Malaya Press.
- Burkhardt, J.L., & Burkhardt, J.M. 2017. Then and now: Changes in social organisation and livelihood of the Berawan community since the formation of Malaysia. *Journal of Borneo-Kalimantan* 3(1), 55-66. <https://doi.org/10.33736/jbk.618.2017>
- Cahill, M., & Karan E. 2008. *Factors in designing effective orthographies for unwritten languages*. SIL International. Retrieved from [https://www.sil.org/system/files/reapdata/71/22/66/71226648004357014393230040163130285698/silewp2008\\_001.pdf](https://www.sil.org/system/files/reapdata/71/22/66/71226648004357014393230040163130285698/silewp2008_001.pdf)
- Ismail Dahaman. 2000. *Pedoman ejaan dan sebutan Bahasa Melayu*. Dewan Bahasa dan Pustaka.
- Merriam-Webster dictionary (n.d.). Economical. In *Merriam-Webster.com dictionary*. Retrieved January 20, 2025, from: <https://www.merriam-webster.com/dictionary/economical>
- Moe, R. 2001. Lexicography and mass production. *Notes on linguistics* 4(3), pp. 150-156.
- Eberhard, D.M., Simons, G. F., & Fennig, C.D. (Eds.). 2024. *Ethnologue: Languages of the world*. (27<sup>th</sup> ed.). SIL International. Online version: <https://www.ethnologue.com>
- Ziu (24 June 2023). The breve - Beyond its use in Romanian. *The Language Closet*. Retrieved January 20, 2025, from: <https://thelanguagecloset.com/2023/06/24/breve/>