

PENGELOMPOKAN BAHASA MELAYU DENGAN BAHASA BANJAR DALAM SALASILAH HIPOTETIKAL BAHASA: KAJIAN PERBANDINGAN LEKSIKOSTATISTIK

MOHD KHAIDIR BIN ABDUL WAHAB
Universiti Sains Malaysia
mohdkhaidir@usm.my

ABSTRAK

Sesebuah bahasa dipencarkan daripada bahasa induk yang pernah wujud sebelum ini. Pemencaran tersebut mewujudkan salasilah bahasa. Dalam salasilah bahasa, filum bahasa merupakan kelompok bahasa induk yang paling tua usianya, iaitu 5,000 tahun ke atas. Rumpun bahasa pula ialah kelompok bahasa yang berusia antara 2,500 hingga 5,000 tahun. Dalam salasilah hipotetikal bahasa, keluarga bahasa ialah kelompok bahasa yang berusia kurang daripada 2,500 tahun. Keluarga bahasa menurunkan subkeluarga bahasa yang masing-masing berlainan usia. Unsur linguistik bahasa mungkin kekal atau berinovasi sewaktu dua bahasa atau lebih dipencarkan daripada bahasa induk. Faktor kesamaan linguistik ini menunjukkan bahawa wujud hubungan kekerabatan antara bahasa turunan atau bahasa yang dibandingkan daripada induk yang sama. Kajian ini meneliti salasilah bahasa antara bahasa Melayu dan bahasa Banjar dengan menyenaraikan kosa kata asas kedua-dua bahasa melalui kaedah perbandingan leksikostatistik. Tinjauan dan penyenaian kosa kata asas bahasa ini memanfaatkan senarai 100 kosa kata Morris Swadesh yang diubah suai. Kaedah ini menentukan peratusan kata berkerabat dengan menghitung jumlah kata berkognat melalui penelitian unsur retensi dan unsur inovasi. Dengan penelitian tersebut, salasilah bahasa bagi kedua-dua bahasa ini dapat ditentukan. Hasil kajian mendapati bahawa daripada 100 perkataan yang dibandingkan, sebanyak 72 kosa kata atau 76.5% bahasa Melayu adalah berkognat dengan bahasa Banjar. Berdasarkan peratus kekerabatan ini, ia memperlihatkan bahawa bahasa Melayu dan bahasa Banjar dikelompokkan pada tingkat kekerabatan keluarga bahasa. Dalam salasilah hipotetikal bahasa, keluarga bahasa ialah kelompok bahasa yang berusia kurang daripada 2,500 tahun. Oleh itu, kajian ini dapat memberi implikasi yang signifikan terhadap persoalan pengelompokan bahasa-bahasa Austronesia berdasarkan salasilah hipotetikal bahasa secara berpada dengan menggunakan kaedah perbandingan leksikostatistik.

Kata Kunci: bahasa Banjar; bahasa Melayu; kosa kata; leksikostatistik; salasilah bahasa

GROUPING MALAY WITH BANJAR LANGUAGE IN THE HYPOTHETICAL GENEALOGY OF LANGUAGE: A COMPARATIVE LEXICOSTATISTICAL STUDY

ABSTRACT

A language is dispersed from a proto language that used to exist before. Such dispersal forms the language genealogy. In language genealogy, the language phylum is the oldest group of proto language, which is 5,000 years old and above. Language clusters are language groups aging between 2,500 to 5,000 years old. In the hypothetical language genealogy, language family consists of language groups that age less than 2,500 years old. Language family descends into language subfamilies with different ages. The linguistic elements of a language may remain or innovate when two or more languages are dispersed from the proto language. This factor of linguistic similarity suggests a relationship between derived languages or languages compared from the same proto language. This study examines the language genealogy between Malay and Banjar by listing the basic vocabulary of both languages through lexicostatistical comparison method. This basic vocabulary survey and listing will make use of the modified list of 100 Morris Swadesh vocabularies. This method determines the percentage of related words by calculating the number of cognate words. The method is carried out by studying the elements of retention and innovation of the language. Therefore, the language genealogy of these two languages can be determined. The results of the study found that out of 100 words compared, 72 vocabularies or 76.5% of the Malay language are cognate with the Banjar language. Based on this percentage, it shows that Malay and Banjar are grouped at the cognates level of the language family. In the hypothetical genealogy of languages, language families are language groups that age less than 2,500 years old. Therefore, this study possesses significant implications in issues of grouping Austronesian languages based on hypothetical language genealogy appropriately by using lexicostatistical comparison methods.

Keywords: Banjar language; Malay language; vocabulary; lexicostatistics; language genealogy

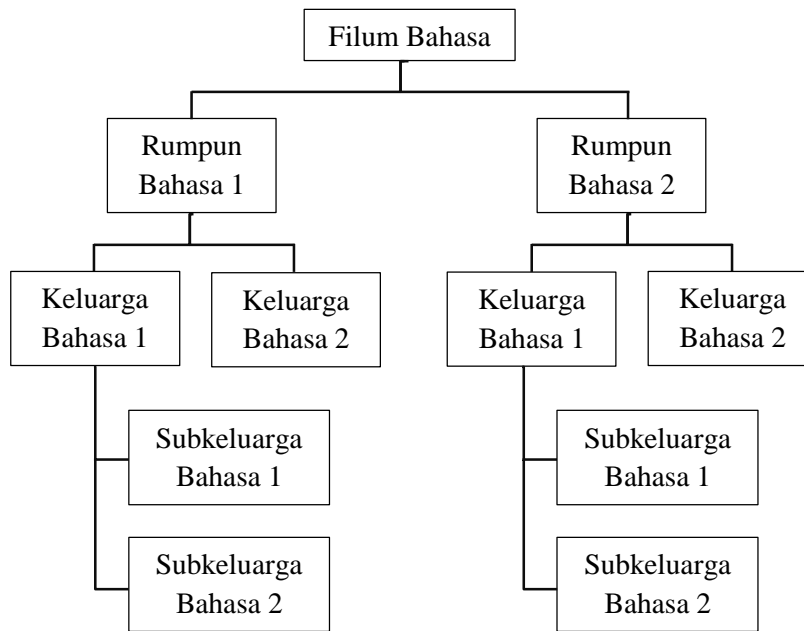
PENGENALAN

Dalam salasilah keluarga manusia, ada hubungan dekat dan jauh. Hubungan dekat diwarisi daripada ibu bapa yang sama iaitu hubungan adik-beradik; abang, kakak dan adik. Hubungan yang jauh sedikit pula ialah hubungan sepupu, dua pupu, tiga pupu dan seterusnya untuk merujuk kepada pemeringkatan dalam salasilah (Mohd Khaidir & Noriah, 2021:23). Perkaitan kekeluargaan antara bahasa-bahasa boleh dianalogikan kepada perkaitan kekeluargaan antara manusia kerana dalam bahasa juga, ada perkaitan yang dekat dan jauh. Sekumpulan bahasa yang memperlihatkan lebih banyak ciri-ciri yang sama antara mereka akan menganggotai satu kelompok berlainan dengan kelompok lain yang mempunyai anggota-anggotanya sendiri, walaupun antara dua kelompok tersebut masih terdapat perkaitan kekeluargaan (Asmah,1993:32).

Salasilah Bahasa

Salasilah bahasa ialah daftar keturunan atau asal usul sesuatu bahasa (Syed Zainal Arif & Noriah, 2017:1). Ini bermaksud, sesuatu bahasa dipencarkan daripada satu bahasa induk yang lebih tua lalu mewujudkan salasilah bahasa. Salasilah bahasa mempunyai peringkat perkembangan bahasa yang tertentu berdasarkan usia bahasa atau istilahnya “lama waktu” yang berbeza-beza. Berdasarkan perkaitan kekeluargaan ini, maka bahasa-bahasa dimasukkan ke dalam kelompok-kelompok yang diberi nama-nama khusus berdasarkan lama-waktu (*time-depth*) masing-masing. Lama-waktu sesuatu bahasa merupakan perkiraan umur bahasa berkenaan yang dikira dari masa bahasa itu dipencarkan daripada induknya. Berdasarkan lama-waktu bahasa-bahasa yang mempunyai perkaitan kekeluargaan, bahasa-bahasa digolongkan ke dalam rumpun, keluarga dan subkeluarga. (Asmah,1993:33).

Konsep “lama waktu” ialah pemeringkatan kepada filum, rumpun, keluarga dan bahasa. Peringkat paling atas dalam hierarki salasilah bahasa ialah filum bahasa yang berusia 5,000 tahun ke atas. Filum bahasa ialah kelompok bahasa induk yang paling tua usia atau “lama waktu”. Filum bahasa menurunkan kelompok rumpun bahasa yang berusia antara 2,500 hingga 5,000 tahun. Rumpun bahasa pula menurunkan kelompok keluarga bahasa yang berusia kurang daripada 2,500 tahun. Akhirnya, keluarga bahasa pula menurunkan subkeluarga bahasa yang berlainan usia masing-masing. Pembahagian rumpun kepada keluarga adalah berdasarkan persamaan dan perbezaan ciri. Apabila sejumlah bahasa mempunyai persamaan ciri yang tinggi, ia akan dikelompokkan dalam kumpulan yang sama. Menurut Asmah (1993:33), ahli linguistik hanya mengelompokkan sesuatu bahasa setakat subkeluarga sahaja. Ini kerana, jika proses ini diteruskan kepada perincian yang lebih kecil iaitu subsubkeluarga, ia akan menjadi lebih rumit. Rajah 1 menunjukkan salasilah hipotetikal bahasa berdasarkan usia.



RAJAH 1. Salasilah Hipotetikal Bahasa Berdasarkan Usia

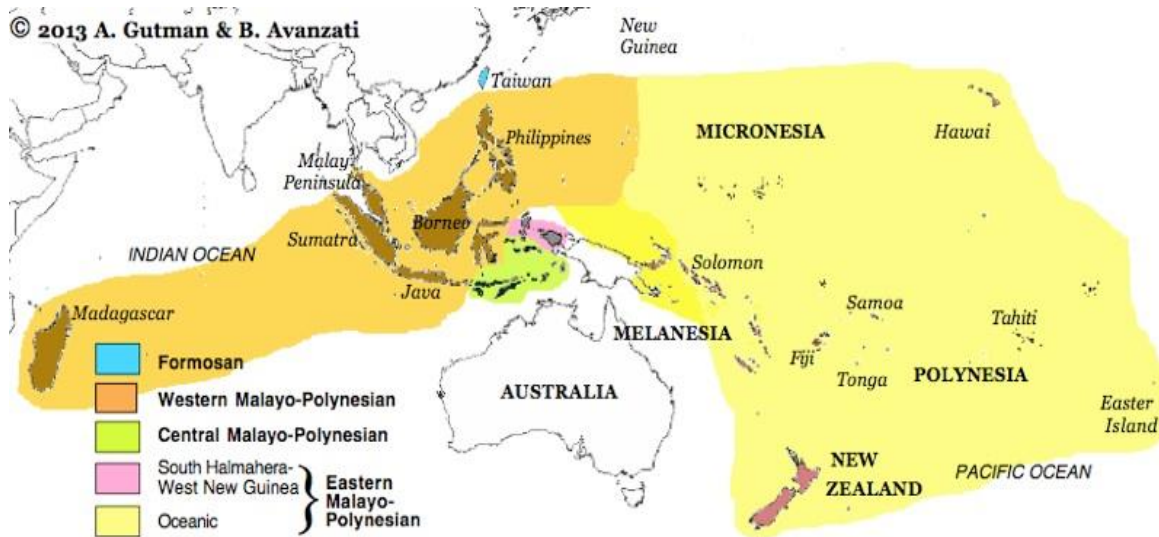
Sumber: Syed Zainal Arif & Noriah (2017:2)

Pemencaran bahasa daripada induk melahirkan bahasa turunan yang akan mewujudkan salasilah bahasa. Salasilah bahasa mengandungi peringkat perkembangan bahasa yang ditentukan berdasarkan usia bahasa atau istilahnya lama-waktu yang berbeza-beza. Perkara ini sejajar dengan salah satu kepentingan bidang linguistik sejarah dan perbandingan (LSP) iaitu untuk mengadakan pengelompokan atau *subgrouping* ke atas bahasa-bahasa yang berada dalam suatu rumpun bahasa atau subkelompok bahasa tertentu.

Kegiatan penghasilan semula bahasa purba atau bahasa induk dianggap sah dan relevan melalui andaian salasilah sesebuah keluarga. Ini kerana, ciri-ciri bahasa purba diwarisi oleh bahasa dan dialek semasa sama ada ciri berkenaan sudah berubah atau masih kekal (Noriah & Rohani, 2011: 5).

Bahasa Austronesia

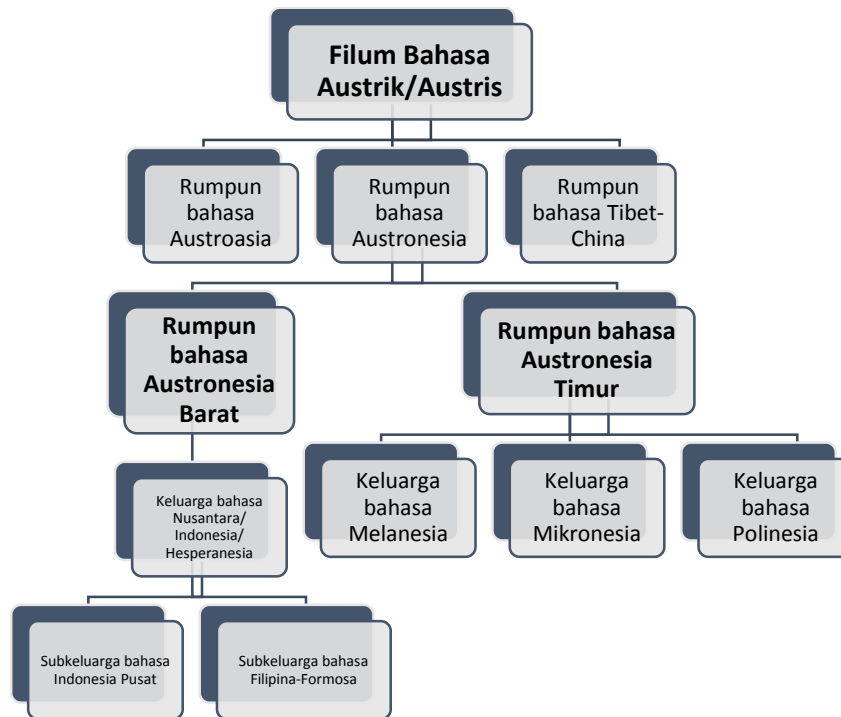
Bahasa Austronesia juga dikenali sebagai bahasa Melayu-Polinesia yang digunakan oleh Wilhelm Von Humbolt dan ahli antropologi. Istilah Bahasa Austronesia dan Melayu Polinesia digunakan untuk merujuk kepada bahasa-bahasa yang terdapat di kawasan Nusantara. Rajah 2 merupakan peta Bahasa Austronesia.



RAJAH 2. Peta Bahasa Austronesia

Sumber: <http://www.languagesgulper.com/>

Bahasa-bahasa rumpun Austronesia tersebar luas dari Formosa di sebelah utara hingga New Zealand di sebelah selatan dan dari Madagaskar di sebelah barat hingga ke Kepulauan Easter di sebelah timur. Di tanah asal Austronesia, diramalkan sudah ada kepelbagaian bahasa dan dialek yang dipencarkan daripada induk Austronesia tersebut. Apabila berlaku perpindahan keluar kelompok orang Austronesia dari kawasan tanah asal, mereka akan membawa bersama bahasa atau dialek mereka (Noriah & Rohani, 2011:18). Dengan adanya persamaan kosa kata antara bahasa di daratan Asia dan kosa kata di pulau-pulau di selatan, ia menyebabkan Wilhelm Von Humbolt membuat satu justifikasi bahawa bahasa-bahasa tersebut berasal daripada induk yang sama. Hubungan kekerabatan itu mengambil masa yang sangat lama, iaitu 5000 tahun ke atas. Induk yang menurunkan bahasa-bahasa di daratan dan di pulau-pulau itu diberi nama filum atau filum Austrik.



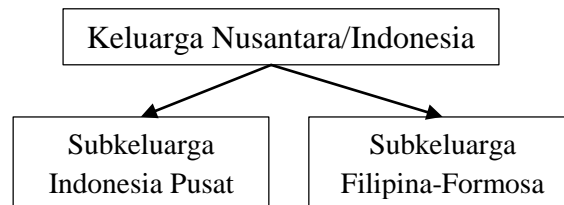
RAJAH 3. Pengelompokan Bahasa-Bahasa Austronesia

Sumber: Syed Zainal Arif & Noriah (2017:3)

Pemencaran bahasa Austronesia purba atau induk kepada bahasa-bahasa atau dialek-dialek turunan yang terdapat di kawasan Austronesia sekarang ini juga dapat dijelaskan menggunakan pendekatan LSP sebagaimana huraian berkaitan salasilah hipotetikal bahasa berdasarkan usia. Berdasarkan andaian asas pendekatan LSP, Filum Austrik atau Austris memencarkan bahasa-bahasa dalam kelompok rumpun Austronesia dan rumpun Austroasia. Perkara ini dapat dilihat pada Rajah 3 di atas.

Istilah Austrik berasal daripada kata dasar dalam bahasa Latin ‘austro’ yang bermakna “selatan”. Justeru, filum Austrik bermakna filum bahasa-bahasa di kawasan selatan. Oleh itu, rumpun Austroasia bermakna “Asia Selatan”. Rumpun Austroasia terdiri daripada bahasa-bahasa Cham dan Mon-Khmer yang semuanya terdapat di dataran Asia, bahagian selatan dan sebahagian besar bahasa orang asli Malaysia. Manakala, istilah Austronesia pula berasal daripada bahasa Latin ‘austro’ dan bahasa Yunani ‘nesos’ yang bermaksud pulau. Dengan itu, “Austronesia” bermakna “pulau-pulau selatan”. Bahasa Melayu dan bahasa orang Asli, iaitu Jakun dan Temuan, digolongkan sebagai bahasa Austronesia. Rumpun Austronesia terbahagi kepada dua kelompok, iaitu rumpun Austronesia Barat dan rumpun Austronesia Timur. Kelompok rumpun Austronesia Barat terdiri daripada keluarga Nusantara. Sementara itu, rumpun Austronesia Timur terdiri daripada tiga keluarga bahasa, iaitu keluarga Melanesia, Mikronesia, dan Polinesia. Pembahagian keluarga-keluarga itu adalah berdasarkan kepada ciri-ciri persamaan dan perbezaan antara satu sama lain. Persamaan dan perbezaan antara keluarga-keluarga itu merupakan persamaan dan perbezaan yang meliputi sistem bunyi, pembentukan kata dan ayat, serta perbendaharaan kata. (Asmah, 1993:33-35).

Keluarga Melanesia berpusat di New guinea (tidak termasuk bahasa-bahasa Papua), Halmahera dan pulau-pulau di sekitarnya. Kawasan Mikronesia pula terdapat di sebelah timur Filipina, iaitu meliputi pulau-pulau Palauan, Chamoro, Guam, Truk dan pulau-pulau lain di sekitarnya. Seterusnya, keluarga Polinesia meliputi bahasa-bahasa yang dituturkan di Fiji dan sekitarnya, dan kawasannya terbentang ke selatan hingga ke New Zealand, dan ke timur hingga ke Hawaii dan Kepulauan Easter. Keluarga Indonesia/Nusantara pula meliputi dari Formosa di utara Kepulauan Filipina dan Maluku yang sempadannya terletak di sebelah timur, dan Pulau Timor di selatan. Di sebelah barat pula ialah Madagaskar.



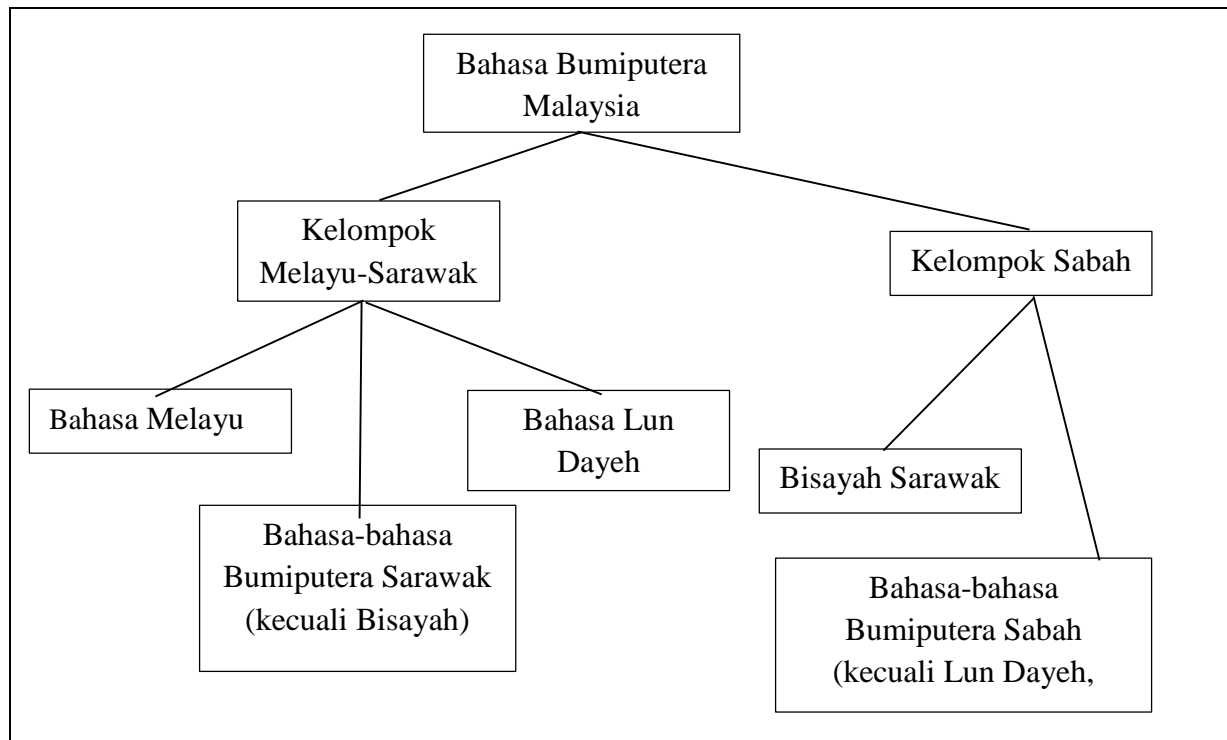
RAJAH 4. Pengelompokan Bahasa-bahasa Austronesia Keluarga Nusantara/Indonesia

Sumber: Noriah & Rohani (2011:20)

Berdasarkan Rajah 4, Keluarga Nusantara/Indonesia boleh dibahagikan kepada subkeluarga Indonesia Pusat dan subkeluarga Filipina-Formosa. Subkeluarga Nusantara/Indonesia Pusat terdiri daripada bahasa Melayu, serta bahasa-bahasa di Sarawak, Kalimantan dan pulau-pulau Indonesia lainnya. Subkeluarga Filipina-Formosa pula terdiri daripada bahasa-bahasa di Filipina, serta bahasa-bahasa di Sabah dan Brunei (kecuali bahasa Melayu dan Lun Dayeh).

Pengelompokan Bahasa Melayu

Bahasa Melayu termasuk dalam subkeluarga Indonesia Pusat. Berdasarkan Rajah 5 di bawah, bahasa Melayu mempunyai perkaitan dengan bahasa-bahasa di Sarawak. Bahasa-bahasa bumiputera di Sabah pula walaupun mempunyai perkaitan yang erat dengan bahasa Melayu, tetapi perkaitannya itu agak jauh sedikit daripada perkaitan bahasa Melayu dengan bahasa-bahasa di Sarawak.



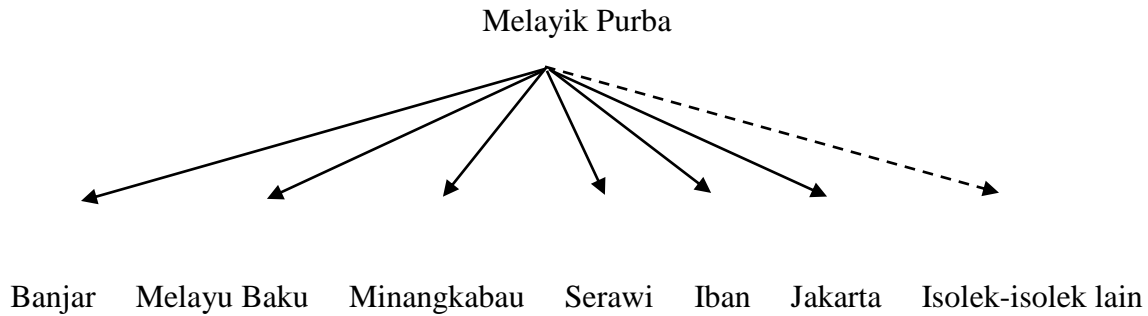
RAJAH 5. Pengelompokan Bahasa Bumiputera Malaysia

Sumber: Asmah (1993:38)

Pengelompokan Bahasa Banjar

Bahasa Banjar merupakan bahasa daerah yang dituturkan oleh suku Banjar, iaitu suatu suku di sekitar Kalimantan Selatan sebagai bahasa ibunda. Penggunaan bahasa ini sangat meluas sehingga meliputi wilayah Kalimantan Selatan dan Kalimantan Timur, Indonesia. Banjar pula merujuk kepada *Bandjarese*, *Banjar Malay* atau *Banjarese* yang dikelompokkan dalam rumpun bahasa Melayik (Rahim Aman *et al.*, 2018: 2).

Subkelompok Melayik Purba terbahagi kepada bahasa Banjar, bahasa Melayu, bahasa Minangkabau, bahasa Iban, bahasa Serawi, bahasa Melayu Jakarta dan beberapa dialek atau bahasa yang ada di pedalaman Kalimantan. Pembahagian subkelompok Melayik adalah seperti berikut:



RAJAH 6. Pembahagian subkelompok bahasa Melayik Purba

Sumber: Adelaar (2004)

Hipotesis bahasa Banjar sebagai bahasa Melayik adalah seperti bahasa Iban, bahasa Minangkabau dan lain-lain. Hipotesis ini dikemukakan kerana taraf bahasa Banjar sebagai *lingua franca*, iaitu bilangan penutur bahasa Banjar semakin bertambah serta lebih banyak daripada bilangan suku Banjar yang sedia ada (Anton M., 1994: 323).

Penghijrahan suku Banjar ke Semenanjung Malaysia telah lama berlaku tetapi tarikh sebenar kedatangan mereka tidak dapat dikenal pasti. Hal ini demikian kerana tiada rekod atau catatan bertulis mengenai penghijrahan mereka. Terdapat dua faktor yang menyebabkan penghijrahan suku Banjar iaitu faktor penolakan dan faktor penarikan. Antara faktor penolakan termasuklah politik, ekonomi dan sosial di Tanah Banjar, Kalimantan yang tidak stabil. Oleh itu, kestabilan pertanian dan perdagangan di Tanah Melayu menjadi faktor penarikan masyarakat Banjar untuk berhijrah (Nur Hafizah & Normaliza, 2016: 56).

Masyarakat Banjar di Malaysia hanya terdapat di beberapa kawasan tertentu iaitu di negeri Johor, Selangor, Sabah dan Perak. Kebanyakan taburan masyarakat Banjar terdapat di beberapa bahagian pantai barat seperti di daerah Kerian iaitu di kawasan Bagan Serai, Parit Buntar, Sungai Manik dan Bagan Datoh di Perak, Sabak Bernam dan Tanjong Karang di Selangor serta Batu Pahat di Johor yang mana kawasan ini memfokuskan kepada penanaman padi. Masyarakat Banjar di Sabah pula terdapat di sekitar daerah Sandakan (Aiza & Norazlan Hadi, 2020: 40; Yusmawati & Mohd Khaidir, 2021:7-8).

Secara amnya, semua bahasa di dunia khususnya di kawasan tertentu mempunyai tingkat kekerabatan masing-masing termasuklah bahasa Melayu dan bahasa Banjar. Sebelum sampai kepada bentuknya sekarang, sudah pasti kedua-dua bahasa ini telah mengalami evolusi sejarah yang panjang setelah berpisah atau dipencarkan daripada bahasa induknya. Justeru, kajian ini akan menentukan pengelompokan kedua-dua bahasa austronesia ini berdasarkan Salasilah Hipotetikal Bahasa dengan menggunakan kaedah leksikostatistik. Seterusnya hubungan kekerabatan jauh atau dekat kedua-dua bahasa ini dapat diketahui secara berpada.

OBJEKTIF KAJIAN

Menentukan pengelompokan bahasa Melayu dengan bahasa Banjar dalam Salasilah Hipotetikal Bahasa dengan membandingkan 100 kosa kata dasar.

METODOLOGI KAJIAN

Kajian ini menggunakan pendekatan kualitatif iaitu dengan cara pengkaji menjalankan kajian lapangan. Temu bual berstruktur telah dilakukan bersama informan masyarakat banjar di mukim Bagan Serai, Perak.



RAJAH 7. Batu Tanda Mukim Bagan Serai, Daerah Kerian, Perak

Lokasi kajian dan Pemilihan Informan

Dalam kajian ini, informan yang dipilih merupakan penutur natif bahasa Banjar di kampung yang dikaji. Mereka tinggal di Kampung Masjid Tinggi, mukim Bagan Serai, Daerah Kerian, Perak Darul Ridzuan. Kampung Masjid Tinggi dalam mukim Bagan Serai telah dipilih dalam kajian ini sebagai sumber utama untuk mendapatkan maklumat. Subkampung yang dipilih ialah Jalan Banjar, Kampung Masjid Tinggi.



RAJAH 8. Papan Tanda Kampung Masjid Tinggi

Informan terdiri daripada lima orang warga emas dan lima orang dewasa yang merupakan penutur asli bahasa Banjar. Pemilihan 10 orang responden ini berdasarkan kepada kemampuan mereka dalam menuturkan bahasa Banjar dengan baik dan telah menetap di kampung tersebut sejak lahir atau dalam tempoh yang agak lama. Informan yang dipilih juga tidak mempunyai masalah pendengaran dan alat artikulasinya berada dalam keadaan baik semasa proses penyebutan kosa kata dasar.

Instrumen Kajian

Pengkaji menggunakan senarai 100 kosa kata dasar bahasa Melayu dengan bahasa Banjar yang telah diubahsuai berasaskan cadangan Swadesh (1955:133-137).

Kaedah Perbandingan

Bidang perbandingan merupakan bidang untuk mempelajari data daripada suatu bahasa atau lebih, sekurang-kurangnya dalam dua tempoh masa. Data sesuatu bahasa dalam dua tempoh masa atau lebih akan dibandingkan secara terperinci bagi memperoleh perubahan yang berlaku dalam bahasa itu (Gorys Keraf, 1984: 22). Kaedah perbandingan pula merujuk kepada bagaimana dua bahasa atau lebih dibandingkan untuk mengungkap kekerabatan antara bahasa pada zaman prasejarah iaitu sebelum wujudnya bahan-bahan bertulis (Noriah, 2001: 380). Bahasa semasa yang wujud zaman kini digunakan untuk mengenal pasti kekerabatan yang dipencarkan daripada bahasa induk atau proto. Ini kerana, dialek mahupun bahasa turunan yang lebih muda berkemungkinan berevolusi mengikut peredaran zaman. Perubahan yang berlaku itu mungkin tidak sama dengan bahasa induk. Oleh hal yang demikian, perbandingan secara khusus dilakukan bagi meneliti inovasi yang berlaku terhadap dialek atau bahasa yang dipencar daripada bahasa induk sekali gus memberi gambaran tentang bahasa induk bagi sesebuah bahasa atau dialek. Namun begitu, perbandingan bahasa dalam kajian ini bukan bertujuan untuk memperoleh gambaran bahasa induk tetapi lebih kepada usaha bagi meneliti perubahan dua bahasa iaitu bahasa Melayu dan bahasa Banjar. Dengan penelitian tersebut, maka pengelompokan kedua-dua bahasa ini dapat ditentukan berdasarkan Salasilah Hipotetikal Bahasa.

Walau bagaimanapun, keupayaan menganalisis menggunakan kaedah kualitatif atau kaedah pemeriksaan unsur bahasa yang kekal (*reliks*) dan berinovasi perlu dilakukan terlebih dahulu untuk

mengenal pasti kosa kata yang berkognat (+), kosa kata yang tidak berkognat (-) dan kosa kata yang perlu disingkirkan.

Kaedah Perbandingan Kualitatif

Kaedah perbandingan kualitatif merujuk kepada pemeriksaan unsur bahasa yang tidak berubah dan berinovasi. Kaedah yang dilaksanakan ini akan mengenal pasti dan membandingkan kosa kata berkognat dalam bahasa-bahasa yang dianggap dalam kelompok yang sama.

Kosa kata berkognat merujuk kepada kosa kata yang sama ataupun mirip sama ada dari segi bentuk mahupun makna dalam bahasa-bahasa yang dibandingkan. Kosa kata tersebut terbina disebabkan oleh kosa kata berkenaan dipencarkan daripada bahasa induk atau proto yang sama. Sebagai contoh, kosa kata kaki **kaki* dalam bahasa Austronesia purba yang disebut sebagai kaki dalam bahasa Melayu dan bahasa Banjar merupakan kosa kata berkognat yang sama bentuk. Manakala, kata **udang* dalam bahasa Austronesia purba yang disebut sebagai *udang* dalam bahasa Melayu dan *hundang* dalam bahasa Banjar pula merupakan kosa kata berkognat yang mirip. Oleh itu, melalui kaedah perbandingan kualitatif, kosa kata berkognat daripada bahasa-bahasa yang dibandingkan akan dikumpulkan. Ini kerana, kosa kata tersebut merupakan kosa kata dasar atau kosa kata budaya. Disebabkan itu, kosa kata yang dibandingkan perlu memperlihatkan kesepadanan.

Selepas bentuk bahasa yang kekal dan berinovasi ditentukan, bahasa-bahasa tersebut dikelompokkan berdasarkan data yang memperlihatkan kesepadanan yang lebih mirip bagi setiap kelompok. Dengan kata lain, jika dua atau lebih daripada dua bahasa memperlihatkan ciri-ciri kekal dan inovasi yang sama, maka bahasa-bahasa berkenaan tergolong dalam kelompok yang sama (Noriah, 2001: 382).

Kaedah Perbandingan Leksikostatistik

Leksikostatistik ialah kaedah yang digunakan untuk menentukan tarikh atau usia kewujudan sesuatu bahasa dari sudut penemuan arkeologi (Gudschinsky, 1956: 1). Noriah dan Ismail (2021:71) pula menyatakan bahawa leksikostatistik ialah kaedah untuk menentukan kadar peratusan kosa kata berkerabat dengan menghitung jumlah kosa kata berkognat bagi meneliti unsur retensi dan unsur inovasi. Perbezaan kaedah leksikostatistik dengan kaedah kualitatif dapat dilihat dari segi jenis kosa kata yang digunakan untuk perbandingan di mana kaedah leksikostatistik menggunakan kosa kata dasar (*basic vocabulary*) sebagai asas penghitungan. Oleh hal yang demikian, melalui kaedah ini, tingkat kekerabatan sesuatu bahasa dapat ditentukan berdasarkan peratus kata berkerabat yang diperolehi.

PENGUMPULAN DATA

Daftar 100 kosa kata dasar bahasa Melayu dan bahasa Banjar seperti Jadual 1 di bawah.

JADUAL 1. Perbandingan 100 Kosa Kata Dasar Bahasa Melayu dengan Bahasa Banjar

Bil	Ejaan		Sebutan	
	Bahasa Banjar	Bahasa Melayu	Bahasa Banjar	Bahasa Melayu
1	can	mengejek	[acan]	[məŋəcəʔ]
2	baayun	bergayut	[ba:yun]	[bərgayut]
3	hundang	udang	[hundəŋ]	[udəŋ]
4	karudut	berkedut	[karudut]	[bəkədut]
5	taburahai	terburai	[taburahai]	[tərburai]
6	umai	amboi	[umai]	[amboi]
7	sigar	segar	[sigar]	[səgar]
8	bigi	biji	[bigi]	[biji]
9	tajun	terjun	[tajun]	[tərjun]
10	anum	muda	[anom]	[muda]
11	lamak	gemuk	[lamaʔ]	[gəmuʔ]
12	caluk	seluk	[caluk]	[səluʔ]
13	jubung	penuh	[jubəŋ]	[pənuh]
14	macal	nakal	[macal]	[nakal]
15	ikam	kamu	[ikam]	[kamu]
16	pisit	ketat	[pisit]	[kətət]
17	layur	layur	[layor]	[layur]
18	hangkup	hantuk	[haŋkəp]	[hantuʔ]
19	ambung	lambung	[ambəŋ]	[ləmbuŋ]
20	saumuran	seumur hidup	[saumuran]	[səumur hidup]
21	kadap	gelap	[kadap]	[kadap]
22	taguh	teguh	[tagoh]	[təguh]
23	ungut	termenung	[uŋot]	[tərmənuŋ]
24	sarik	marah	[sarəʔ]	[marah]
25	layau	merayau	[layau]	[mərayau]
26	maigau	mengigau	[maigau]	[məŋigau]
27	injing	Tarik	[injəŋ]	[tariʔ]
28	tabarurut	terurut	[tabarurut]	[təruurut]
29	nginum	minum	[ŋinum]	[minum]
30	babau	berbau	[babau]	[bərbau]
31	puruk	pakai	[purəʔ]	[pakai]
32	simbur	simbah	[simbor]	[simbah]
33	ampar	hampar	[ampar]	[hampar]
34	rakai	rosak	[rakai]	[rosaʔ]
35	ungal	bergerak	[uŋal]	[bərgərəʔ]
36	tumbus	tembus	[tumbus]	[təmbus]
37	catuk	ketuk	[catəʔ]	[kətəʔ]
38	muntung	mulut	[muntuŋ]	[mulut]
39	tataguk	burung hantu	[tataguʔ]	[buruŋ hantu]
40	wihang	rahang	[wihaŋ]	[rahaŋ]

41	lamari	almari	[lamari]	[almari]
42	tuha	tua	[tuha]	[tua]
43	marinaan	anak saudara	[marinaʔan]	[anaʔ saudara]
44	babat	ikat	[babat]	[ikat]
45	kancang	kencang	[kancaŋ]	[kəncəŋ]
46	olah	buat	[olah]	[buat]
47	putik	petik	[putɛʔ]	[pətiʔ]
48	culup	celup	[culɔp]	[cəlɔp]
49	adding	adik	[adiŋ]	[adiʔ]
50	napang	kenapa	[napaŋ]	[kənapa]
51	rumbih	tanah runtuh	[rumbih]	[tanah runtuh]
52	hayam	ayam	[hayam]	[ayam]
53	takutan	takut	[takutan]	[takut]
54	urang	orang	[uraŋ]	[oraŋ]
55	balar	calar	[balar]	[calar]
56	amun	jika	[amon]	[jika]
57	muhara	depan	[muhara]	[dəpan]
58	pacul	tanggal	[pacul]	[taŋgal]
59	egal	tari	[ɛgal]	[tari]
60	cibuk	ceduk	[cibuʔ]	[cəduʔ]
61	anggap	ambil hati	[aŋgap]	[ambil hati]
62	wani	berani	[wani]	[bərani]
63	talam	dulang	[talam]	[dulaŋ]
64	dadai	jemur	[dadai]	[jəmur]
65	lawas	lama	[lawas]	[lama]
66	masigit	masjid	[masigit]	[masjid]
67	bangat	sangat	[baŋat]	[saŋat]
68	saruan	undangan	[saruan]	[undaŋan]
69	kambang	kembang	[kambaŋ]	[kəmbaŋ]
70	pander	berbual	[pander]	[bərbual]
71	tajarungkup	tersungkur	[tajaruŋkup]	[tərsuŋkur]
72	bahalolong	melolong	[bahaloloŋ]	[məloloŋ]
73	ganal	besar	[ganal]	[bəsar]
74	lali	lupa	[lali]	[lupa]
75	tabuk	tebuk	[taboʔ]	[təboʔ]
76	sumap	kukus	[sumap]	[kukus]
77	katuju	setuju	[katuju]	[sətuju]
78	guha	gua	[guha]	[gua]
79	parak	dekat	[paraʔ]	[dəkaʔ]
80	yat	lihat	[yat]	[lihat]
81	warang	besan	[waraŋ]	[bəsan]
82	batatukar	membeli	[batutukar]	[məmbəli]
83	tiharap	tiarap	[tiharap]	[tiarap]
84	hatap	atap	[hatap]	[atap]
85	bagamat	perlahan	[bagamat]	[pərlahan]
86	katup	tutup	[katɔp]	[tutɔp]
87	rawa	sapa	[rawa]	[sapa]
88	tihang	tiang	[tihaŋ]	[tiaŋ]

89	tatawa	ketawa	[tatawa]	[kətawa]
90	sasah	kejar	[sasah]	[kəjar]
91	pancau	tinggi	[pancau]	[tiŋgi]
92	batingas	berkelahi	[batiŋas]	[bərkelahi]
93	saing	rajin	[saɛŋ]	[rajin]
94	wayah	zaman	[wayah]	[zaman]
95	tuhuk	teruk	[tuhuʔ]	[təruʔ]
96	uyuh	lelah	[uyuh]	[ləlaʔ]
97	haruk	busuk	[haroʔ]	[busuʔ]
98	tabarusuk	terperosok	[tabarusoʔ]	[tərpərosoʔ]
99	galah	halau	[galah]	[halau]
100	arum	harum	[arom]	[harum]

PENGANALISISAN DATA

Langkah Perbandingan Kosa Kata Dasar Menggunakan Kaedah Leksikostatistik

Perbandingan kosa kata dasar dalam leksikostatistik didahului dengan meneliti kosa kata dasar bagi bahasa Melayu dan bahasa Banjar. Langkah-langkah penelitian yang perlu diikuti adalah seperti berikut. Pertama, menentukan beberapa kosa kata yang digunakan dan kosa kata yang perlu ditolak. Kosa kata yang patut ditolak ialah a) kata pinjaman (misalnya, ‘*almari*’ dalam bahasa Melayu atau “*lamari*” dalam bahasa Banjar yang dipinjam daripada bahasa Portugis); b) kata kosong (ada dalam satu bahasa sahaja); dan c) kata majmuk (contoh dalam perbandingan ini ‘burung hantu’ atau ‘tataguk’). Kedua, jika terdapat kata terbitan, imbuhan perlu disingkirkan dan hanya kata dasar sahaja diambil. Namun begitu, jika kata dasar tersebut sudah wujud, maka kata dasar selepasnya perlu ditolak. Ketiga, apabila semua kata yang tidak berkenaan telah selesai ditolak, barulah jumlah kata boleh ditetapkan untuk dibandingkan. Keempat, membandingkan kosa kata dan menentukan taraf kognat antara kosa kata.

Untuk menentukan sama ada pasangan kosa kata itu berkognat ataupun tidak, kriteria dalam perbandingan kualitatif seperti yang berikut diambil kira, iaitu a) pasangan kata itu mempunyai persamaan bentuk dan makna secara menyeluruh (*total*); dan b) pasangan kata itu memiliki bentuk dan makna yang sepadan atau mirip. Hasil perbandingan akan memperlihatkan jumlah kata berkognat antara kedua-dua bahasa (Noriah dan Ismail, 2021:73).

Peratus kata berkerabat, iaitu *c* diperoleh dengan membahagikan jumlah kata berkognat dengan jumlah kata yang dibandingkan iaitu *n*. Jumlah kata yang dibandingkan ini merujuk kepada jumlah kosa kata dasar selepas ditolak kosa kata dasar yang tidak diperlukan iaitu kata pinjaman, kata kosong dan kata majmuk.

Berdasarkan peratusan kata berkerabat, pengelompokan bahasa berdasarkan Salasilah Hipotetikal Bahasa dapat ditentukan melalui penggunaan jadual tingkat kekerabatan bahasa seperti yang ditunjukkan di dalam Jadual 2.

JADUAL 2. Jadual Tingkat Kekerabatan Bahasa

Tingkat kekerabatan	Peratus kata berkerabat
Bahasa	100 – 81
Keluarga	81 – 36
Rumpun	36 – 12
Mikrofilum	12 – 4
Mesofilum	4 – 1
Makrofilum	Kurang daripada 1

[Sumber: Dipetik dan diubah suai daripada Gudschinsky,1964: 621]

DAPATAN KAJIAN

Jadual 3 berikut merupakan perbandingan antara taraf kognat atau tidak berkognat pada kosa kata dasar bahasa Melayu dengan bahasa Banjar. Sebanyak 100 kosa kata direkodkan di dalam Jadual tersebut.

JADUAL 3. Perbandingan Taraf Kognat atau Tidak Berkognat 100 Kosa Kata Dasar Bahasa Melayu dengan Bahasa Banjar.

Bil	Ejaan		Sebutan		Taraf Kognat (+) atau (-)
	Bahasa Banjar	Bahasa Melayu	Bahasa Banjar	Bahasa Melayu	
1	can	mengejek	[acan]	[məŋəcəʔ]	-
2	baayun	bergayut	[ba:yun]	[bərgayut]	+
3	hundang	udang	[hundəŋ]	[udəŋ]	+
4	karudut	berkedut	[karudut]	[bəkədut]	+
5	taburahi	terburai	[taburahi]	[təburai]	+
6	umai	amboi	[umai]	[amboi]	+
7	sigar	segar	[sigar]	[səgar]	+
8	bigi	biji	[bigi]	[biji]	+
9	tajun	terjun	[tajun]	[tərun]	+
10	anum	muda	[anom]	[muda]	-
11	lamak	gemuk	[lamaʔ]	[gəmuʔ]	+
12	caluk	seluk	[caluk]	[səluʔ]	+
13	jubung	penuh	[jubəŋ]	[pənuh]	-
14	macal	nakal	[macal]	[nakal]	+
15	ikam	kamu	[ikam]	[kamu]	+
16	pisit	ketat	[pisit]	[kətət]	+
17	layur	layur	[layor]	[layur]	+
18	hangkup	hantuk	[həŋkəp]	[hantuʔ]	+

19	ambung	lambung	[ambonʝ]	[lambuŋ]	+
20	saumuran	seumur hidup	[saumuran]	[səumur hidup]	TOLAK
21	kadap	gelap	[kadap]	[kadap]	+
22	taguh	teguh	[tagoh]	[təguh]	+
23	ungut	termenung	[uŋot]	[tərmənunʝ]	-
24	sarik	marah	[sarɛʔ]	[marah]	-
25	layau	merayau	[layau]	[mərayau]	+
26	maigau	mengigau	[maigau]	[məŋigau]	+
27	injing	tarik	[injen]	[tariʔ]	-
28	tabarurut	terurut	[tabarurut]	[tərorut]	+
29	nginum	minum	[ŋinum]	[minum]	+
30	babau	berbau	[babau]	[bərbau]	+
31	puruk	pakai	[purɔʔ]	[pakai]	+
32	simbur	simbah	[simbor]	[simbah]	+
33	ampar	hampar	[ampar]	[hampar]	+
34	rakai	rosak	[rakai]	[rosaʔ]	+
35	ungal	bergerak	[uŋal]	[bərgərəʔ]	-
36	tumbus	tembus	[tumbus]	[təmbus]	+
37	catuk	ketuk	[catɔʔ]	[kətɔʔ]	+
38	muntung	mulut	[muntunʝ]	[mulut]	+
39	tatagak	burung hantu	[tataguʔ]	[burunʝ hantu]	TOLAK
40	wihang	rahang	[wihanʝ]	[rahanʝ]	+
41	lamari	almari	[lamari]	[almari]	TOLAK
42	tuha	tua	[tuha]	[tua]	+
43	marinaan	anak saudara	[marinaʔan]	[anaʔ saudara]	TOLAK
44	babat	ikat	[babat]	[ikat]	+
45	kancang	kencang	[kancanʝ]	[kəncanʝ]	+
46	olah	buat	[olah]	[buat]	-
47	putik	petik	[putɛʔ]	[pətiʔ]	+
48	culup	celup	[culɔp]	[cəlɔp]	+
49	adding	adik	[adiŋ]	[adiʔ]	+
50	napang	kenapa	[napanʝ]	[kənapa]	-
51	rumbih	tanah runtuh	[rumbih]	[tanah runtuh]	TOLAK
52	hayam	ayam	[hayam]	[ayam]	+
53	takutan	takut	[takutan]	[takut]	+
54	urang	orang	[uranʝ]	[oranʝ]	+
55	balar	calar	[balar]	[calar]	+
56	amun	jika	[amon]	[jika]	-
57	muhara	depan	[muhara]	[dəpan]	-
58	pacul	tanggal	[pacul]	[tanʝal]	-
59	egal	tari	[egal]	[tari]	-
60	cibuk	ceduk	[cibuʔ]	[cəduʔ]	+
61	anggap	ambil hati	[anʝap]	[ambil hati]	TOLAK
62	wani	berani	[wani]	[bərani]	+
63	talam	dulang	[talam]	[dulənʝ]	+
64	dadai	jemur	[dadai]	[jəmʝur]	+
65	lawas	lama	[lawas]	[lama]	+
66	masigit	masjid	[masigit]	[masʝid]	+

67	bangat	sangat	[baŋat]	[saŋat]	+
68	saruan	undangan	[saruan]	[undaŋan]	+
69	kambang	kembang	[kambaŋ]	[kəmbaŋ]	+
70	pander	berbual	[paŋdər]	[bərbual]	-
71	tajarungkup	tersungkur	[tajaruŋkup]	[tərsuŋkur]	+
72	bahalolong	melolong	[bahalolon]	[məlonon]	+
73	ganal	besar	[ganal]	[bəsar]	-
74	lali	lupa	[lali]	[lupa]	+
75	tabuk	tebuk	[taboʔ]	[təboʔ]	+
76	sumap	kukus	[sumap]	[kukus]	-
77	katuju	setuju	[katuju]	[sətuju]	+
78	guha	gua	[guha]	[gua]	+
79	parak	dekat	[paraʔ]	[dəkat]	-
80	yat	lihat	[yat]	[lihat]	+
81	warang	besan	[waraŋ]	[bəsan]	-
82	batatukar	membeli	[batutukar]	[məmbəli]	-
83	tiharap	tiarap	[tiharap]	[tiarap]	+
84	hatap	atap	[hatap]	[atap]	+
85	bagamat	perlahan	[bagamat]	[pərlahan]	-
86	katup	tutup	[katəp]	[tutəp]	+
87	rawa	sapa	[rawa]	[sapa]	+
88	tihang	tiang	[tihaŋ]	[tiaŋ]	+
89	tatawa	ketawa	[tatawa]	[kətawa]	+
90	sasah	kejar	[sasah]	[kəjar]	-
91	pancau	tinggi	[pancau]	[tiŋgi]	-
92	batingas	berkelahi	[batiŋas]	[bərkəlahi]	+
93	saing	rajin	[saɛŋ]	[rajin]	+
94	wayah	zaman	[wayah]	[zaman]	+
95	tuhuk	teruk	[tuɦuʔ]	[təruʔ]	+
96	uyuh	lelah	[uyuh]	[ləlaʔ]	+
97	haruk	busuk	[haroʔ]	[busuʔ]	+
98	tabarusuk	terperosok	[tabarusəʔ]	[tərpərosoʔ]	+
99	galah	halau	[galah]	[halau]	+
100	arum	harum	[arom]	[harum]	+

Berdasarkan jadual ini, terdapat enam kosa kata dasar yang ditolak dengan alasan masing-masing iaitu kosa kata yang terdapat pada bilangan 20,31,43,51 dan 61 kerana kata majmuk. Manakala, kosa kata pada bilangan 41 ialah kata pinjaman.

Selepas ditolak satu kosa kata ini, maka jumlah kosa kata yang dibandingkan dapat ditentukan, iaitu $n = 100 - 6 = 94$. Seterusnya, semua kosa kata dibandingkan dan ditentukan taraf kognat dengan menggunakan kriteria perbandingan kualitatif. Kosa kata berkognat diberi tanda campur (+), manakala yang tidak berkognat bertanda tolak (-). Berdasarkan Jadual 3, peratus kosa kata berkerabat antara bahasa Melayu dan bahasa Banjar adalah seperti berikut:

Jumlah kosa kata yang ditolak = 6

Jumlah kosa kata yang dibandingkan (100-6)	= 94
Jumlah kosa kata yang berkognat (94-22)	= 72
PERATUS KOSA KATA BERKOGNAT (72/94)	= 0.765
	= 76.5%

Jadual 4. Jadual Tingkat Keperabatan Bahasa Banjar dengan Bahasa Melayu.

Tingkat kekerabatan	Peratus kata berkerabat
Bahasa	100 – 81
Keluarga	81 – 36
Rumpun	36 – 12
Mikrofilum	12 – 4
Mesofilum	4 – 1
Makrofilum	Kurang daripada 1

Sumber: Dipetik dan diubahsuai daripada Gudschinsky (1964: 621)

Berdasarkan dapatan kajian melalui perbandingan dua buah bahasa Austronesia iaitu bahasa Melayu dengan bahasa Banjar melalui kaedah leksikostatistik, hasil menunjukkan peratus kata berkerabat adalah antara 36 hingga 81 dan dikelompokkan pada tingkat kekerabatan keluarga bahasa dalam Salasilah Hipotetikal Bahasa, iaitu kelompok bahasa yang berusia kurang daripada 2,500 tahun.

KESIMPULAN

Kajian ini telah membincangkan dan menerapkan kaedah perbandingan leksikostatistik antara dua bahasa Austronesia iaitu bahasa Melayu dan bahasa Banjar. Daripada 100 perkataan yang dibandingkan sebanyak 72 kosa kata atau 76.5% bahasa Melayu yang berkognat dengan bahasa Banjar. Berdasarkan peratus kekerabatan ini memperlihatkan bahasa Melayu dan bahasa Banjar dikelompokkan pada tingkat kekerabatan keluarga bahasa. Dalam salasilah hipotetikal bahasa, keluarga bahasa ialah kelompok bahasa yang berusia kurang daripada 2,500 tahun. Kajian ini merupakan suatu usaha kecil pengkaji dalam mendokumentasikan kosa kata bahasa Banjar dan hubungan kekerabatan dengan bahasa Melayu yang dipencarkan daripada induk yang sama. Di samping itu, kajian ini juga dapat membuktikan bahawa kaedah leksikostatistik adalah relevan diguna pakai dalam menghuraikan salasilah hipotetikal bahasa berdasarkan usia yang mewujudkan istilah kekeluargaan bahasa seperti filum, rumpun, keluarga, bahasa, dialek dan seterusnya secara berpadu.

RUJUKAN

- Adelaar, K. 2004. "Where does Malay come from? Twenty Years of Discussions about Homeland, Migrations and Classifications" dalam *Bijdragen tot de Taal-, Land-en Volkenkunde (160) (1)*, 1-30.
- Aiza Maslan @ Baharuddin dan Norazian Hadi Yaacob. 2020. Komuniti Banjar di Tanah Rantau: Ibadah Haji sebagai Satu Dorongan Tradisi Masyarakat Banjar "Madam Ka Banua Urang". PPIK USM, Jabatan Kemasyarakatan dan Kewarganegaraan UPSI. *Jurnal Perspektif*, 12, 39 – 52.
- Asmah Omar. 1993. *Susur Galur Bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Anton M. Moeliono. 1994. *Bahasawan Cendiaika*. West Java: Fakultas Sastra, Universitas Indonesia.
- Gorys, K., 1984. *Linguistik Bandingan Historis*. Jakarta: Penerbit PT Gramedia Pustaka Utama.
- Gudschinsky, S.C. 1956. The ABC'S of lexicostatistics (glottochronology). *WORD*, 12(2), 175-210.
- Mohd Khaidir Abdul Wahab & Noriah Mohamed. 2021. Hubungan Sejarah dan Perkaitan Linguistik Antara Bahasa dalam Noriah Mohamed (Peny.) *Menelusuri Intisari Linguistik Sejarah dan Perbandingan*. Kuala Lumpur: Institut Terjemahan & Buku Malaysia.
- Nasrun Alias & James T. Collins. 2017. Pendokumentasian Dialek Ulu Terengganu Sebagai Wahana Komunikasi: Satu Tinjauan Awal Aspek Morfologi. *Jurnal Melayu*, Isu Khas, 80-91.
- Norfazila Ab. Hamid & Rahim Aman. 2016. Varian Melanau Sarawak: Tinjauan di Melanau Mukah. *Jurnal Melayu*, 15 (1), 99-112.
- Norfazila Ab Hamid. 2020. Varian Melanau Sarawak: Tinjauan di Melanau Igan. *Jurnal Melayu*. 19 (2), 321-336.
- Norhaniza Abdul Khairi, Shahidi A. Hamid & Rahim Aman. 2021. Realisasi Kontras Penyuaaran Bunyi Plosif Awal Kata Dialek Melayu Patani Lenggong. *Jurnal Melayu*, 20(1), 215-223.
- Noriah Mohamed & Ismail Mohd. 2021. Kesamaan Linguistik dan Kaedah Perbandingan Kuantitatif dalam Noriah Mohamed (Peny.) *Menelusuri Intisari Linguistik Sejarah dan Perbandingan*. Kuala Lumpur: Institut Terjemahan & Buku Malaysia.

- Noriah Mohamed dan Rohani Mohd. Yusof. 2011. *Pengelompokan Leksikostatistik Bahasa-Bahasa Peribumi di Sarawak*. Pulau Pinang: Universiti Sains Malaysia
- Noriah Mohamed. 2001. Perbandingan Kuantitatif Bahasa Miri dengan Bahasa Melayu. *Jurnal Bahasa (1)(3)*, 378 – 397.
- Nur Hafizah Razali & Normaliza Abd Rahim. (2016). Makna Ujaran Masyarakat Banjar dari Sudut Pragmatik. *Journal of Business and Social Development*, 4(1), 55 – 64.
- Rahim Aman, Nazihah Najwa Othman, Shahidi A. H, Ery Iswary & Anwar O.D. 2018. Rekonstruksi Dalaman Varian Banjar Beriah Purba. *Jurnal of Nusantara Studies*, 3(2), 92 – 106
- Swadesh, M. 1955. *Towards greater accuracy in lexicostatistic dating*. *International Journal of American Linguistics*, 21(2), 121-137. Diakses daripada <https://sci-hub.tw/10.1086/464321>.
- Syed Zainal Ariff Syed Jamaluddin & Noriah Mohamed. 2017. *Susur Galur dan Perkembangan Imbuhan Bahasa Melayu Klasik*. Selangor: PTS Akademia.
- Yusmawati Mat Yusof dan Mohd Khaidir Abdul wahab (2021). Kata Sapaan dan Rujukan dalam kalangan Masyarakat Banjar Terpilih di Daerah Kerian, Perak. *International Journal Of Language Education and Applied Linguistics (IJLEAL)*, 11 (2), 6-22.

Biodata Penulis:

Mohd Khaidir bin Abdul Wahab (PhD) merupakan Pensyarah Kanan di Bahagian Pengajian Linguistik Bahasa Melayu, Pusat Pengajian Ilmu Kemanusiaan, Universiti Sains Malaysia. Bidang kepakaran penulis ialah sejarah bahasa Melayu. Beliau juga memfokuskan kajian dalam bidang linguistik sejarah bahasa-bahasa Austronesia.