

Kajian Leksikostatistik Pada Bahasa Batak Toba Dan Batak Angkola Mandailing

oleh

Lilis Handayani Napitupulu¹, lilishandayaninapitupulu@unprimdn.ac.id

Yolanda Novita Silaban², yolandanovitasilaban@unprimdn.ac.id

Universitas Prima Indonesia

ABSTRAK-Penelitian ini bertujuan untuk mendeskripsikan tingkat kekerabatan, masa pisah, dan tahun prediksi bahasa Batak Toba dan bahasa Batak Angkola Mandailing berdasarkan persamaan dan perbedaan fonologinya dengan menggunakan teori linguistik historis komparatif. Dalam pengumpulan data digunakan metode kepustakaan. Kemudian dalam mengkaji data digunakan metode leksikostatistik. Hasil penelitian menunjukkan bahwa kekerabatan bahasa Batak Toba dengan Batak Angkola Mandailing bahwa: (1) dari 100 kosakata Swadesh hanya 55 kata yang berkerabat dengan kategori 45 kategori tidak berkerabat. Tingkat kekerabatan kedua bahasa mempunyai hubungan kekerabatan pada tingkat keluarga (family) dengan persentase 55%, (2) bahasa Batak Toba dan Batak Angkola Mandailing merupakan bahasa tunggal pada 118 – 1334 tahun yang lalu dengan perhitungan $1377 + 43 = 118$ tahun dan $1377 - 43 = 1334$ tahun dan mulai berpisah sekitar tahun 640 Masehi ($2017 - 1377 = 640$).

Kata kunci: leksikostatistik, bahasa batak toba, bahasa batak angkola mandailing, linguistik historis komparatif

A. Pendahuluan

Linguistik merupakan satu ilmu yang otonom dan mempergunakan metode-metode ilmiah. Dikatakan linguistik merupakan sebuah ilmu empiris, dengan metode induktif. Dalam studi bahasa sebagai ilmu sering diberdakan pula atas linguistik dan fonetik. Para linguis Inggris benar-benar membedakan studi linguistik dan studi fonetik, sebagai dua disiplin ilmu yang berbeda. Dikatakan bahwa bidang fonetik begitu penting khususnya dalam sistem ejaan dan ucapan bahasa Inggris, sehingga sukar bagi mereka untuk memasukkannya dalam kerangka besar studi linguistik. Dikatakan bahwa fonetik dapat dipelajari lepas dari semua unsur bahasa-bahasa yang lain.

Fonetik dan linguistik mempelajari aspek yang berbeda dari

fenomena yang sama. Fonetik mempelajari suara, sedangkan linguistik mempelajari organisasinya. M.A.K. Halliday, seorang linguis Inggris menambahkan bahwa dua bidang itu disatukan dalam satu teori yang umum yang disebut linguistik umum. Linguistik umum (*linguistic general*) merumuskan dan memberikan batasan mengenai obyek studinya. Apa yang dirumuskan oleh linguistik umum harus mengenai semua bahasa manusia yang bersifat alamiah dan menentukan bagaimana semua diuraikan dan dibandingkan. Linguistik umum berisikan sebuah teori umum mengenai bahasa dan bertujuan menelusuri dan memberikan batas kepada kategori-kategori yang umum dari gejala bahasa dan mempelajari serta menentukan apa yang sistematis dalam ujaran dan bahasa.

Pembahasan dibatasi hanya dalam satu bahasa daerah saja yakni bahasa Batak. Etnik Batak terdiri atas beberapa sub etnik yaitu Toba, Simalungun, Karo, Pakpak, Dairi, dan Angkola Mandailing. Kelima sub etnik ini memiliki bahasa yang satu sama lain mempunyai banyak persamaan. Tetapi disini dibedakan sedikitnya dua cabang bahasa Batak Toba dan Mandailing untuk melihat kekerabatan bahasa, masa pisah bahasa, dan sekaligus prediksi usia bahasa antara kedua cabang bahasa Batak tersebut dengan menggunakan kajian leksikostatistik.

Leksikostatistik merupakan bagian linguistik historis komparatif yakni bidang linguistik yang menyelidiki perbandingan suatu bahasa lain (Ridwan, 1995:3). Linguistik historis komparatif dapat menentukan hubungan kekerabatan antar bahasa-bahasa yang seasal. Disamping istilah leksikostatistik ada juga istilah lain yaitu glotokronologi yang kenyataannya kedua istilah tersebut saling melengkapi, maka sering pula keduanya disamakan. Kajian glotokronologi lebih mengutamakan perhitungan waktu.

Penelitian merupakan suatu usaha untuk mengumpulkan data atau fakta, serta pelaksanaan konsep untuk mencari dan memperoleh atau mendapatkan kebenaran yang sanggup mengambil lebih dalam kebenaran yang sudah ada. Dalam hal ini adapun tujuan penelitian yaitu untuk mengetahui kekerabatan antara bahasa Batak Toba dengan Batak Angkola Mandailing yang ditinjau berdasarkan kosa kata dasar, masa pisah antara bahasa Batak Toba dengan bahasa Batak Angkola Mandailing, prediksi usia bahasa Batak Toba dengan bahasa Batak Angkola Mandailing.

Salah satu metode pengelompokan adalah leksikostatistik. Disamping istilah leksikostatistik dikenal juga istilah lain yaitu glotokronologi. Pada dasarnya kedua istilah tersebut sebenarnya memiliki pengertian yang agak berlainan, sekurang-kurangnya menyangkut sasaran akhir yang akan dicapai. Pengertian pokok antara kedua istilah itu dapat dirumuskan sebagai berikut:

1. Leksikostatistik adalah suatu teknik dalam pengelompokan bahasa yang lebih cenderung mengutamakan peneropongan kata-kata (leksikon) secara statistik, untuk kemudian berusaha menetapkan pengelompokan itu berdasarkan presentase kesamaan dan perbedaan suatu bahasa dengan bahasa lain.
2. Glotokronologi adalah suatu teknik dalam linguistik historis yang berusaha mengadakan pengelompokan dengan lebih mengutamakan perhitungan waktu atau perhitungan usia bahasa-bahasa kerabat. Dalam hal ini usia bahasa tidak dihitung secara mutlak dalam suatu tahun tertentu, tetapi dihitung secara umum, misalnya mempergunakan satuan ribuan tahun.

Dalam kenyataan kedua bidang itu selalu dipakai secara bergandengan, karena untuk menghitung usia bahasa dengan teknik glotokronologi harus dipergunakan leksikostatistik. Sebaliknya untuk mengadakan pengelompokan bahasa dengan metode leksikostatistik tersirat juga masalah waktu, yang menjadikan landasan bagi pengelompokan bahasa-bahasa berlandaskan usia bahasa, kedua istilah itu dipakai dengan pengertian yang sama.

Dengan demikian, leksikostatistik atau glotokronologi dapat dibatasi

sebagai suatu teknik yang berusaha menemukan keterangan-keterangan (data-data) untuk suatu tingkat waktu yang agak tua dalam bahasa guna menentukan usia bahasa dan pengelompokan bahasa-bahasa.

Dari penelitian-penelitian yang dilakukan terhadap pelbagai bahasa, akhirnya diperoleh empat macam asumsi dasar yang dapat dipergunakan sebagai titik tolak dalam usaha mencari jawaban mengenai usia bahasa, atau secara tepatnya bilamana terjadi diferensiasi antara dua bahasa atau lebih. Asumsi-asumsi dasar tersebut adalah:

1. Sebagian dari kosa kata suatu bahasa sukar sekali berubah bila dibandingkan dengan bagian lainnya.
2. Retensi (ketahanan) kosa kata dasar adalah konstan sepanjang masa.
3. Perubahan kosa kata dasar pada semua bahasa adalah sama.
4. Bila presentase dari dua bahasa kerabat (cognate) diketahui, maka dapat dihitung waktu pisah kedua bahasa tersebut.

Berdasarkan asumsi dasar yang kedua, ketiga, dan keempat, dapat dihitung usia atau waktu pisah bahasa A dan B kalau diketahui presentase kata kerabat kedua bahasa itu. Dan karena dalam tiap 1.000 tahun kedua bahasa kerabat itu masing-masing akan kehilangan kosa kata dasarnya dalam presentase yang sama, maka waktu pisah antara kedua bahasa itu harus dibagi dua. Misalnya presentase kata kerabatnya adalah 80,5 % maka waktu pisah kedua bahasa adalah 500 tahun yang lalu.

Untuk menerapkan keempatasumsi dasar diatas, maka perlu diambil langkah-langkah tertentu yang sekaligus merupakan teknik-teknik metode leksikostatistik, yaitu:

1. Mengumpulkan kosa kata dasar bahasa kerabat.

Unsur yang paling penting dalam membandingkan dua bahasa atau lebih adalah mengumpulkan daftar kosa kata dasar dari bahasa-bahasa yang akan diteliti. Daftar yang baik adalah daftar yang disusun oleh Morris Swadesh yang berisi 200 kata. Daftar tersebut membawa keuntungan dalam penelitian karena terdiri dari kata-kata yang non-kultural, serta retensi kata dasarnya telah diuji dalam bahasa-bahasa yang memiliki naskah-naskah tertulis.

2. Menetapkan pasangan-pasangan mana dari kedua bahasa tadi adalah kata kerabat (cognate).

Dalam membandingkan kata-kata untuk menetapkan kata-kata mana yang merupakan kata kerabat dan mana yang tidak, maka perlu dikemukakan lagi suatu asumsi lain dalam metode perbandingan yaitu fonem bahasa proto yang sudah berkembang secara berlainan dalam bahasa-bahasa kerabat akan berkembang terus secara konsisten dalam lingkungan linguistik masing-masing bahasa kerabat.

3. Menghitung usia atau waktu pisah kedua bahasa.

Waktu pisah antara dua bahasa kerabat yang telah diketahui presentase kata kerabatnya, dapat dihitung dengan mempergunakan rumus berikut:

$$t = \frac{\log. C}{2 \log. r}$$

Dimana t = waktu perpisahan dalam ribuan tahun yang lalu; r = retensi atau presentase konstan dalam 1000 tahun, atau disebut juga indeks, C = presentase kerabat, log. = logaritma dari.

Rumus diatas dapat diselesaikan dengan mengikuti tahap-tahap berikut:

- a. Mula-mula mencari logaritma C dan r dalam daftar logaritma;
 - b. Kemudian logaritma r dikalikan dengan 2;
 - c. Hasil logaritma C dibagi dengan hasil dari sebelumnya;
 - d. Hasil dari pembagian sebelumnya menunjukkan waktu pisah dalam satuan ribuan tahun.
4. Menghitung jangka kesalahan untuk menetapkan kemungkinan waktu pisah yang lebih tepat.

Cara yang biasa dipergunakan untuk menghindari kesalahan dalam statistik adalah memberi suatu perkiraan bahwa suatu hal terjadi bukan dalam waktu tertentu, tetapi dalam suatu jangka waktu tertentu. Untuk menghitung jangka kesalahan biasanya dipergunakan kesalahan standar, yaitu 70% dari kebenaran yang diperkirakan. Kesalahan standar, yaitu 70% dari kebenaran yang diperkirakan. Kesalahan standar diperhitungkan dengan rumus berikut:

$$S = \sqrt{\frac{C(I-C)}{n}}$$

Dimana S = kesalahan standar dalam presentase kata kerabat; C = presentase kata kerabat; n = jumlah kata yang diperbandingkan (baik kerabat maupun non kerabat). Perhitungannya dapat dilakukan dengan cara mengikuti urutan berikut:

1. I dikurangi C;
2. C dikalikan dengan hasil dari (1);
3. Hasil dari (2) dibagi dengan n;
4. Menarik akar atas hasil dari (3);
5. Hasil dari (4) merupakan jangka kesalahan dari presentase kata kerabat atas dasar 0,7 % perkiraan mengenai kebenaran yang sesungguhnya.

Metode leksikostatistik dan glotokronologi bukan semata-mata merupakan metode untuk menentukan waktu pisah dua bahasa kerabat, tetapi ia juga menjadi metode untuk mengadakan pengelompokan bahasa-bahasa kerabat. Sebenarnya dengan mengetahui waktu pisah antara bahasa-bahasa kerabat berdasarkan kata-kata kerabatnya sudah tersirat pula pengelompokan-pengelompokan itu. Bahasa-bahasa yang memperlihatkan presentase kekerabatan yang tinggi merupakan kelompok yang lebih dekat keanggotaannya sedangkan yang presentase kekerabatannya kecil saja merupakan bahasa yang agak jauh kekerabatannya dan termasuk dalam kelompok yang lebih besar.

Dengan mempergunakan dasar-dasar leksikostatistik, Swadesh mengusulkan suatu klasifikasi untuk menetapkan kapan dua bahasa disebut dialek, kapan sekelompok bahasa disebut keluarga bahasa (*Language of Family*), bilamana sekelompok bahasa termasuk rumpun bahasa (*stock*) dan sebagainya. Klasifikasi yang dimaksud adalah sebagai berikut:

Tingkatan Bahasa	Waktu Pisah Dalam Abad	Presentase Kata Kerabat
Bahasa (<i>Language</i>)	0-5	100-81
Keluarga (<i>Family</i>)	5-25	81-36
Rumpun (<i>Stock</i>)	25-50	36-12
Mikrofilum	50-75	12-4
Mesofilum	75-100	4-1
Makrofilum	100-keatas	1-kurang dari 1

B. Metode Penelitian

Metode adalah cara yang teratur dan terpikir baik untuk mencapai maksud, cara kerja sistematis untuk memudahkan pelaksanaan penelitian guna mencapai tujuan yang ditemukan. Penelitian merupakan rangkaian kegiatan ilmiah dalam rangka pemecahan suatu permasalahan. Jadi, penelitian merupakan bagian dari usaha pemecahan masalah. Fungsi penelitian adalah mencari penjelasan dan jawaban terhadap permasalahan serta memberikan alternatif bagi kemungkinan yang dapat digunakan untuk pemecahan masalah. Penjelasan dan jawaban terhadap permasalahan itu dapat bersifat abstrak dan umum sebagaimana halnya dalam penelitian dasar dan dapat pula sangat konkret dan spesifik seperti biasanya ditemui dalam penelitian terapan. Metode penelitian sendiri adalah langkah dan prosedur yang akan dilakukan dalam pengumpulan data atau informasi guna memecahkan permasalahan dan menguji hipotesis penelitian.

Sesuai dengan masalah yang dikaji, metode yang dipergunakan adalah metode leksikostatistik. Metode leksikostatistik adalah metode pengelompokan bahasa yang dilakukan dengan menghitung persentase perangkat kognat (Mahsun, 1995:115). Kosakata yang menjadi dasar perhitungan adalah kosakata Swadesh. Adapun langkah-langkah dalam metode leksikostatistik antara lain:

1. Mengumpulkan kosakata Swadesh
2. Menghitung kata kerabat
3. Menghitung waktu pisah
4. Menghitung jangka kesalahan

Metode pengumpulan data dilakukan agar dapat memiliki acuan sumber-sumber data yang cukup. Untuk

memperoleh data-data yang dibutuhkan, penulis melakukan studi kepustakaan dengan mencari buku-buku sebagai bahan acuan dari berbagai referensi sehingga lebih mendukung dalam menganalisis masalah.

Adapun prosedur yang harus diikuti sebagai analisis data adalah metode dasar yakni langkah-langkah yang harus ditempuh dalam penerapan asumsi dasar diikuti dengan adanya rumus-rumus leksikostatistik yang telah ditentukan. Adapun prosedur yang harus diikuti sebagai analisis data adalah sebagai berikut:

1. Mengumpulkan kosakata Swadesh
2. Menghitung kata kerabat yakni dengan mengikuti prosedur yang sudah ditentukan seperti:
 - a. Gloss yang tidak diperhitungkan
 - b. Pengisolasian morfem terikat
 - c. Penetapan kata kerabat

Setelah itu untuk menghitung presentase kata kerabat digunakan rumus:

$$C = \frac{k}{n} \times 100\%$$

c: cognates atau kata kerabat
k: jumlah kata yang berkerabat
n: jumlah kata yang dibandingkan

3. Menghitung waktu pisah
Waktu pisah antara kedua bahasa kerabat yang telah diketahui presentase kata kerabatnya, dapat dihitung dengan mempergunakan rumus sebagai berikut:

$$t = \frac{\text{Log } c}{2 \text{ Log } r}$$

t: time depth (waktu pisah)
r: retention (daya tahan kosa kata)
c: cognate (persentase kata kerabat)

4. Menghitung jangka kesalahan
5. Untuk menghitung jangka kesalahan biasanya dipergunakan kesalahan standar, yaitu 70% dari kebenaran yang diperkirakan. Kesalahan standar yang diperhitungkan dengan rumus sebagai berikut:

$$S = \frac{C(I-C)}{n}$$

S: kesalahan standar dalam presentase kata kerabat

C: presentase kata kerabat

n: jumlah kata yang diperbandingkan

Hasil dari kesalahan standar ini dijumlahkan dengan presentase kerabat untuk mendapatkan c baru. Dengan c yang baru ini sekali lagi dihitung waktu pisah dengan mempergunakan rumus waktu pisah pada teknik c. Setelah diperoleh hasil jangka kesalaham, maka waktu yang lama dikurangi dengan waktu yang baru. Angka inilah ditambah dan dikurangi dengan waktu yang lama untuk memperoleh usia atau waktu pisah kedua bahasa itu.

C. Hasil Penelitian dan Pembahasan

Leksikostatistik adalah suatu teknik dalam pengelompokan bahasa yang lebih cenderung mengutamakan peneropongan kata-kata (leksikon) secara statistik, untuk kemudian berusaha menetapkan pengelompokan itu berdasarkan presentase kesamaan dan perbedaan suatu bahasa dengan bahasa lain (Keraf, 1984). Dengan demikian, yang ingin dicapai dalam teknik ini adalah kepastian mengenai usia bahasa, yaitu mengenal kapan sebuah bahasa muncul dan bagaimana hubungannya dengan bahasa-bahasa kerabat lainnya.

Kata kerabat:

$$C = \frac{k}{n} \times 100\%$$

k: jumlah kata yang berkerabat

n: jumlah kata yang dibandingkan

$$C = \frac{k}{n} \times 100\%$$

$$= \frac{100}{61} \times 100\%$$

$$C = 61\%$$

Selanjutnya, dihitung waktu pisah dari bahasa yang dibandingkan, dengan rumus:

$$t = \frac{\text{Log } c}{2 \text{ Log } r}$$

t: time depth (waktu pisah)

r: retention (daya tahan kosa kata)

c: cognate (persentase kata kerabat)

$$2 \text{ Log } r$$

$$= \frac{\text{Log } 61}{2 \log 805}$$

$$= \frac{494}{2 \times 217}$$

$$= \frac{494}{434}$$

$$t = 1,138$$

Ini berarti bahwa kelompok bahasa A (Batak Toba) dan bahasa B (Batak Angkola Mandailing) berpisah 1,138 tahun yang lalu.

Metode penentuan waktu pisah kedua kelompok bahasa tersebut adalah

glotokronologi. Glotokronologi adalah suatu teknik dalam LHK yang berusaha mengelompokkan dua bahasa dengan mengutamakan perhitungan waktu pisah atau perhitungan usia bahasa-bahasa berkerabat. Level pengelompokannya adalah:

Level Subgrouping	Years of separation
Dialects of language	Less than 500 years
Language of family	500 to 2500 years
Family of stock	2500 to 5000 years
Stocks of a microphylum	5000 to 7500 years
Microphyla of a mesophylum	7500 to 10000 years
Mesophyla of a macrophylum	More than 10000 years

Jadi, bahasa Batak Toba dan bahasa Batak Angkola Mandailing termasuk kelompok keluarga bahasa (Language of Family). Bila dihitung dengan tahun sekarang (2017), kedua bahasa tersebut berpisah sekitar tahun 640 Masehi (2017 – 1377 = 640).

Jangka Kesalahan:

$$S = \frac{C(I-C)}{n}$$

S: kesalahan standar dalam presentase kata kerabat
 C: presentase kata kerabat
 n: jumlah kata yang diperbandingkan

Perhitungannya dapat dilakukan dengan mengikuti urutan berikut:

1. I dikurangi C;
2. C dikalikan dengan hasil dari (1);
3. Hasil dari (2) dibagi dengan n;
4. Menarik akar atas hasil dari (3);

5. Hasil dari (4) merupakan jangka kesalahan dari prosentase kata kerabat atas dasar 0.7 perkiraan mengenai kebenaran yang sesungguhnya.

$$\begin{aligned}
 S &= \sqrt{\frac{C(1-C)}{n}} \\
 &= \sqrt{\frac{0,61(1-0,61)}{100}} \\
 &= \sqrt{\frac{0,61 \times 0,39}{100}} \\
 &= \sqrt{\frac{0,22}{100}} \\
 &= \sqrt{0,0022} \\
 &= 0,0469041576
 \end{aligned}$$

S = 0,05

Hasil dari kesalahan standar ini (0,05) dijumlahkan dengan presentase kerabat untuk mendapatkan C baru. Dengan c yang baru ini sekali lagi dihitung waktu pisah dengan mempergunakan rumus waktu pisah pada teknik no. C.

$$\begin{aligned}
 t &= \frac{\text{Log } c}{2 \text{ Log } r} \\
 &= \frac{\text{Log } (0,61 + 0,05)}{434} \\
 &= \frac{\text{Log } 0,66}{434} \\
 &= \frac{416}{434} \\
 &= 0,958525
 \end{aligned}$$

t = 0,958

Untuk memperoleh jangka kesalahan, waktu yang lama dikurangi dengan waktu yang baru ($1138 - 958$) = 180. Angka ini harus ditambah dan dikurangi dengan waktu yang lama untuk memperoleh usia dan waktu pisah kedua bahasa tersebut.

Bahasa Batak Toba dan Bahasa Batak Angkola Mandailing merupakan bahasa Tunggal pada $1138 + 180 = 1318$; $1377 - 180 = 958$.

Bahasa Batak Toba dan Bahasa Batak Angkola Mandailing merupakan bahasa tunggal pada 1318-958 tahun yang lalu.

Bahasa Batak Toba dan Bahasa Batak Angkola Mandailing mulai berpisah dari bahasa protoanya antara 699-1059 sebelum masehi (dihitung dari tahun 2017).

Kesimpulan

Berdasarkan uraian teoritis yang dikemukakan pada leksikostatistik memberi perbandingan antara bahasa Batak Toba dan bahasa Batak Angkola Mandailing maka disimpulkan sebagai berikut:

1. Kelompok bahasa A (Batak Toba) dan bahasa B (Batak Angkola Mandailing) berpisah 1,138 tahun yang lalu.
2. Bahasa Batak Toba dan bahasa Batak Angkola Mandailing termasuk kelompok keluarga bahasa (*Language of Family*).
3. Bila dihitung dengan tahun sekarang (2017), Bahasa Batak Toba dan bahasa Batak Angkola Mandailing berpisah sekitar tahun 879 Masehi ($2017 - 1138 = 879$).
4. Bahasa Batak Toba dan Bahasa Batak Angkola Mandailing merupakan bahasa Tunggal pada $1138 + 180 = 1318$; $1377 - 180 = 958$.

5. Bahasa Batak Toba dan Bahasa Batak Angkola Mandailing merupakan bahasa tunggal pada 1318-958 tahun yang lalu.

6. Bahasa Batak Toba dan Bahasa Batak Angkola Mandailing mulai berpisah dari bahasa protoanya antara 699-1059 sebelum masehi (dihitung dari tahun 2017).

Daftar Pustaka

- Chaer, Abdul. 1994. *Linguistik Umum*. Jakarta: Rineka Cipta.
- Crowley, Terry & Bower, Claire. 2010. *An Introduction to Historical Linguistics*. USA: Oxford University Press, Inc.
- John, Lyons. 1995. *Pengantar Teori Linguistik*. Jakarta: P.T. Gramedia.
- Keraf, Gorys. 1984. *Linguistik Bandingan Historis*. Jakarta: P.T. Gramedia.
- Parera, Jos Daniel. 1986. *Kajian Linguistik Umum Historis Komparatif dan Tipologi Struktural*. Jakarta: P.T. Gelora Aksara Pratama.

Lampiran

No	Kosa Kata Dasar Swadesh	Arti dalam Bahasa Batak Toba	Arti dalam Bahasa Batak Mandailing	A (=) Toba B (=) Mandailing	
				A	B
1	abu	abu	abu	A	A
2	air	aek	aek	A	A
3	akar	akar	akar	A	A
4	alir (me)	ḍorus	alir	A	B
5	anak	dakdanak	anak	A	B
6	angin	alogo	anjɪn	A	B
7	anjing	bianɔ	anjɪŋ	A	B
8	apa	aha	aha	A	A
9	api	apɪ	apɪ	A	A
10	apung (me)	maŋɔmbanɔ	tarapɔŋ	A	B
11	asap	timus	asap	A	B
12	awan	awan	awan	A	A
13	ayah	amanɔ/bapak	aya	A	A
14	bagaimana	səŋɔn dia	bia	A	A

	na				
15	baik	borjo	borjo	A	A
16	bakar	totoŋ	bakar	A	B
17	balik	balik	balik	A	A
18	banyak	godan	bahat	A	B
19	baring	gadap/pagadap	gulan	A	B
20	baru	baru/imbaru	baru	A	A
21	basah	bornok	maræk	A	B
22	batu	bato	bato	A	A
23	beberapa	pigapiga	tarbahat	A	B
24	belah (me)	bola	bola	A	A
25	benar	toto	botol	A	A
26	bengkak	bonkakat	marbosar	A	B
27	benih	boni	bibit	A	A
28	berat	sorat	borat	A	A
29	berenang	marlan	marlan	A	A
30	beri	lehon	lehen	A	A
31	berjalan	mardalan	dalan	A	A
32	besar	balga	godan	A	B
33	bilamana	sadihari	anggo	A	B
34	binatan	binatan	binatan	A	A
35	bintang	bintan	bintan	A	A
36	buah	parboe	boah	A	B
37	bulan	bulan	bulan	A	A
38	bulu	imbulo	bulo	A	A
39	bunga	bona	bona	A	A
40	bunuh	mamono	pamate	A	B
41	buru (ber)	boru	boru	A	A
42	buruk	borok	joko	A	B
43	burung	pidon	onggas	A	B
44	busuk	bankoroson	bosok	A	B
45	cacing	gilok	cacin	A	B
46	cium	omma	omma	A	A
47	cuci	basu	basu	A	A
48	daging	johot	johot	A	A
49	dan	dohot	dohot	A	A
50	danau	tao	danao	A	B
51	darah	modar	modar	A	A
52	datang	ro	ro	A	A
53	daun	bolon	bolon	A	A
54	debu	orbok	abu	A	B
55	dekat	jonok	jonok	A	A
56	dengan	dohot	dohot	A	A
57	dengar	bege	bege	A	A
58	di dalam	bagasan	i bagasan	A	A

59	di mana	didia	i dia	A	A
60	di sini	dison	mdon	A	B
61	di situ	disi	isadon	A	B
62	di, pada	di	di	A	A
63	dingin	borgo/malom/gi	ngali	A	B
64	diri (ber)	jonjon	jonjon	A	A
65	dorong	dorong	tulak	A	B
66	dua	dua	dua	A	A
67	duduk	hondol	jogok	A	B
68	ekor	ihor	ikor	A	A
69	empat	opat	opat	A	A
70	engkau	ho	ho	A	A
71	gali	halu	galu	A	A
72	garam	sira	sira	A	A
73	garuk	garot	garok	A	A
74	gemuk, lemak	mokmok	mokmok	A	A
75	gigi	ipon	ipon	A	A
76	gigit	harat	sargot	A	B
77	gosok	gosok	gosok	A	A
78	gunung	dolak	gonon	A	B
79	hantam	godap/laplap	topar	A	B
80	hapus	sasa	apos	A	B
81	hati	roha	ate	A	B
82	hidung	igon	igon	A	A
83	hidup	holo	holo	A	A
84	hijau	ijo	rata	A	B
85	hisap	isap	iccap	A	A
86	hitam	biron	lomlom	A	B
87	hitung	eton	eton	A	A
88	hujan	udan	udan	A	A
89	hutan	haraŋan	haraŋan	A	A
90	ia	ibana	ia	A	A
91	ibu	omak	omak	A	A
92	ikan	denke	gulaen	A	B
93	ikat	ihot	kobet	A	B
94	ini	nion	mdon	A	A
95	isteri	dolan saripe	adaboro	A	B
96	itu	i	mdo	A	B
97	jahit	jait	jait	A	A
98	jalan (ber)	dalan	mardalan	A	A
99	jantung	posoposo	janton	A	B
100	jatuh	dabu	madabu	A	A

10
0 61