



Jurnal Bahasa Indonesia bagi Penutur Asing (JBIPA)

journalhomepage: ojs.badanbahasa.kemdikbud.go.id/jurnal/index.php/bipa

ISSN 2685-5135 (Print) | ISSN 2685-8053 (Electronic)



Pemanfaatan Linguistik Korpus dalam Menentukan Kata Berfrekuensi Tinggi pada Buku Sahabatku Indonesia BIPA 1

Diyah Wahyuningtyas^{1*}, Tri Mastoyo Jati Kesuma²

*Universitas Gadjah Mada**, *Universitas Gadjah Mada*

Correspondences author: Jl. Sarimulya No. 65, Kota Madiun, 63111, Indonesia;

Email: diyahwahyuningtyas@mail.ugm.ac.id

article info

Article history:

Received 11 November 2021

Revised 22 November 2021

Accepted 27 November 2021

Available online 30 November 2021

Keywords:

Corpus Linguistics; High frequency words;

“Sahabatku Indonesia BIPA 1 Umum

abstract

This research is motivated by the needs of BIPA 1 students for a list of words that they will encounter in the book Sahabatku BIPA 1. The purpose of this study is describing the list of words with the frequency of their use in the book through the use of corpus linguistics. This research was conducted for 1 month (1 to 30 September 2021). This study uses quantitative methods to obtain research data and qualitative methods to analyze and conclude. The results of the study stated that there were 1,623 word types (different words) and 11,176 tokens (the number of word occurrence). The ten words that have the highest frequency is: *dan* (402); *di* (289); *saya* (219); *yang* (199); *ada* (159); *ini* (159); *dengan* (145); *untuk* (125); *kegiatan* (113); *apa* (106). The benefits of the results of this study can be used as material for making a special dictionary for BIPA1 students.

2021 Jurnal Bahasa Indonesia bagi Penutur Asing (JBIPA). This is an open access article under the CC BY-NC license (<https://creativecommons.org/licenses/by-nc/4.0/>)

DOI: <https://doi.org/0.26499/jbipa.v3i2.4125>

Pendahuluan

Pengajaran BIPA (Bahasa Indonesia bagi Penutur Asing) merupakan salah satu pintu untuk menjadikan bahasa Indonesia mendunia. Sebuah pengajaran tak akan terlepas dari bahan ajar yang digunakan. Bahan ajar tersebut berisi topik-topik pilihan yang di dalamnya ada ribuan kata yang digunakan secara berulang-ulang atau bahkan cukup satu kali saja ditemukan. Kata dalam bahan ajar tersebut mengundang rasa penasaran pemelajar terutama mengenai arti/makna kata, sehingga mereka bisa menggunakannya sesuai kebutuhan.

Kata menjadi salah satu bagian yang penting dalam bahan ajar yang digunakan dalam pembelajaran. Johns & Wilke (2018) menyatakan bahwa *words are important* ‘kata-kata itu penting’. Lightbown & Spada (2011) mengungkapkan bahwa *as it has often been remarked, we can communicate by using words that are not placed in the proper order, pronounced perfectly, or marked with the proper grammatical morphemes, but communication often breaks down if we do not use the correct word* ‘seperti yang sering dikatakan, kita dapat berkomunikasi dengan menggunakan kata-kata yang tidak ditempatkan dalam urutan yang tepat, diucapkan dengan sempurna, atau ditandai dengan morfem gramatikal yang tepat, tetapi komunikasi sering terputus jika kita tidak menggunakan kata yang benar’. Berdasarkan pendapat tersebut maka dapat disimpulkan menjadi dua hal mengenai kata yaitu penting dan benar.

Pemahaman kata dengan tingkat kemunculan (frekuensi penggunaan) yang tinggi dalam bahan ajar dapat membantu pemelajar memahami bahan ajar secara mandiri maupun dengan bimbingan. Siagian (2020) menyebutkan bahwa berdasarkan beberapa hasil penelitian tentang peranan kosakata dalam pembelajaran bahasa asing menunjukkan bahwa pemanfaatan kosakata berfrekuensi tinggi (HFW) berperan penting dalam menentukan keberhasilan pembelajaran bahasa asing. Pengajar terkadang berfokus pada teks sesuai topik yang disediakan dalam bahan ajar, sehingga kata berfrekuensi tinggi (yang terdapat dalam bahan ajar) cenderung terlewatkan.

Maka, latar belakang penelitian adalah kata berfrekuensi tinggi dari buku/bahan ajar (Sahabatku Indonesia BIPA 1) yang digunakan sebagai penunjang pembelajaran perlu diketahui dan diberitahukan kepada pemelajar serta pemanfaatan linguistik korpus dalam menentukan kata berfrekuensi tinggi dari buku tersebut. Berdasarkan latar belakang penelitian, rumusan masalahnya adalah bagaimana memanfaatkan linguistik korpus dalam menentukan kata berfrekuensi tinggi dalam buku Sahabatku Indonesia BIPA 1.

A corpus is a collection of texts, written or spoken, which is stored on a computer ‘Korpus adalah kumpulan teks, tertulis atau lisan, yang disimpan di komputer’ (O’Keeffe *et al.*, 2007). Menurut Arum & Winarti, (2020) korpus dikatakan “alami” karena teks yang dikumpulkan adalah teks yang diproduksi dan digunakan secara wajar dan tidak dibuat-buat (apa adanya) seperti buku ajar, buku teks, dan banyak lagi. Berdasarkan penjelasan tersebut, korpus merupakan kumpulan teks (tertulis atau lisan) yang bersifat alami seperti pada buku ajar dan disimpan di komputer.

Crawford & Eniko (2016) menyatakan bahwa *corpus linguistics is concerned with understanding how people use language in various contexts* ‘linguistik korpus berkaitan dengan pemahaman bagaimana orang menggunakan bahasa dalam berbagai konteks’. Adolphs (dalam, Hizbullah *et al.*, 2016) mengungkapkan bahwa ilmu ini (linguistik korpus) secara khusus meneliti bahasa melalui seperangkat data yang bersifat alamiah, riil sesuai penggunaannya, baik itu data tulisan maupun data lisan yang ditranskripsikan. Dalam penelitian ini, linguistik korpus dimanfaatkan untuk menganalisis seperangkat data yaitu seluruh teks tertulis yang didapat dari buku ajar berjudul “Sahabatku Indonesia BIPA 1” secara digital atau menggunakan perangkat lunak (yang dijalankan di komputer).

O’Keeffe *et al.*, (2007) berpendapat bahwa *the basic functions of corpus software, such as generating word frequency lists and also try to give an idea of the wide range of applications of a*

corpus to fields as diverse as language teaching ‘fungsi dasar perangkat lunak korpus, seperti menghasilkan daftar frekuensi kata dan juga mencoba memberikan gambaran tentang berbagai aplikasi korpus ke berbagai bidang seperti pengajaran bahasa’. O’Keeffe *et al.*, (2007) mengungkapkan bahwa *common corpus technique that software can perform is very fast counting of a word frequency list (or word list) for any set of texts* ‘teknik korpus umum yang dapat dilakukan perangkat lunak adalah penghitungan yang sangat cepat dari daftar frekuensi kata (atau daftar kata) untuk kumpulan teks apa pun’. Berdasarkan penjelasan tersebut, salah satu fungsi dasar perangkat lunak korpus adalah menghasilkan frekuensi daftar kata dalam teks yang ditentukan (dihitung) dengan sangat cepat.

Crawford & Eniko (2016) menyebutkan bahwa Laurence Anthony mengembangkan program perangkat lunak yang sangat berguna untuk analisis linguistik korpus dan sampai saat ini tersedia untuk PC dan Mac di sini: www.laurenceanthony.net/software.html. Salah satu program perangkat lunak tersebut adalah *AntConc* untuk menganalisis leksikal serta gramatikal. Crawford & Eniko (2016) menambahkan bahwa *AntConc* mampu memfasilitasi beberapa jenis analisis seperti *KWIC (keyword in context)* ‘kata kunci dalam konteks’, *n-grams* ‘n-gram’, *collocates* ‘kolokasi’, dan *word lists* ‘daftar kata’. Analisis *word lists* ‘daftar kata’ dapat membuat daftar kata sederhana dengan jumlah frekuensi dan peringkat dari kumpulan data Anda sendiri (Crawford & Eniko, 2016). Bahkan, Anthony (2004) menyebutkan bahwa *AntConc hosts a comprehensive set of tools including a powerful concordancer, word and keyword frequency generators, tools for cluster and lexical bundle analysis, and a word distribution plot* ‘*AntConc* menyelenggarakan seperangkat alat yang komprehensif termasuk concordancer yang kuat, generator frekuensi kata dan kata kunci, alat untuk analisis kumpulan klaster dan leksikal, dan plot distribusi kata.

Arum & Winarti (2020) menyatakan bahwa analisis frekuensi memungkinkan peneliti mengenali kata-kata yang paling sering muncul di korpus tertentu, dan kemudian membandingkan serta membedakannya dengan kata-kata lain. Johns & Wilke (2018) menambahkan bahwa *a list of 100-200 high-frequency words will make up over 50% of the words in school texts* ‘daftar 100-200 kata berfrekuensi tinggi akan membentuk lebih dari 50% kata dalam teks sekolah/buku teks’. Bahkan Johns & Wilke (2018) mengungkapkan bahwa *students who know a core of 200 or so high-frequency words by sight will possess a solid foundation for reading* ‘siswa yang mengetahui inti dari 200 atau lebih kata berfrekuensi tinggi dengan penglihatan akan memiliki dasar yang kuat untuk membaca’. Dapat disimpulkan bahwa *high-frequency words* ‘kata berfrekuensi tinggi’ sebaiknya disampaikan kepada pemelajar bahasa khususnya bahasa asing (*foreign language*) maupun bahasa kedua (*second language*).

Kata berfrekuensi tinggi dalam penelitian ini bersumber dari buku “Sahabatku Indonesia BIPA 1”. Jenjang BIPA 1 yang merupakan bagian dari BIPA pemula. Menurut Siagian (2020) BIPA pemula (1 & 2) merupakan level dengan jumlah pemelajar yang paling tinggi. Sebagai lembaga pemerintahan, Badan Pengembangan Bahasa dan Perbukuan (Kementerian Pendidikan dan Kebudayaan), berperan dengan memberikan fasilitas bagi lembaga penyelenggara program BIPA, pengajar, dan pemelajar BIPA. Salah satu bagian dari fasilitasi tersebut adalah penyediaan bahan ajar.

Pada tahun 2019, telah dihasilkan bahan ajar acuan dan bahan pendukung pembelajaran BIPA yang dikelompokkan ke dalam empat seri bahan, salah satunya adalah bahan ajar BIPA yang ditujukan bagi pemelajar BIPA kategori umum tanpa batasan usia dengan judul “Sahabatku Indonesia”. Buku “Sahabatku Indonesia BIPA 1 Umum” memiliki lingkup kompetensi yaitu pemelajar mampu memahami dan menggunakan ungkapan konteks pengenalan diri dan pemenuhan kebutuhan konkret sehari-hari dan rutin dengan cara sederhana untuk berkomunikasi dengan mitra tutur yang sangat kooperatif. Bahan ajar ini tersusun dalam 10 unit dan tiap unit terdiri dari empat kegiatan berbahasa (menyimak, berbicara, membaca, dan menulis), pengetahuan kebahasaan (tata bahasa), serta wawasan keindonesiaan (informasi tentang budaya Indonesia).

Penelitian sebelumnya mengenai pemanfaatan linguistik korpus menunjukkan bahwa terdapat 12.395 kata berbeda yang digunakan dengan total jumlah kata secara keseluruhan termasuk kata yang diulang-ulang adalah sebanyak 428.117 kata (Arum & Winarti, 2020). Menurut Arum & Winarti (2020), kata-kata tersebut akan digunakan untuk menyusun bahan ajar ESP bidang radiologi dan sebuah kamus mini buku teks radiologi tersebut.

Penelitian lainnya adalah mengenai kata berfrekuensi tinggi (*high frequency word/HFW*) pada buku BIPA pemula yang menyebutkan bahwa 1) HFW bahasa Indonesia untuk pemelajar BIPA berbeda dengan HFW bahasa Indonesia secara umum; 2) bentuk kelas kata yang diajarkan bervariasi; dan 3) bentuk imbuhan yang digunakan terbatas (Siagian, 2020). Menurut Siagian (2020), pemanfaatan HFW dalam pembelajaran bahasa asing terbukti memberi efek positif, seperti meningkatkan motivasi belajar, meningkatkan rasa percaya diri untuk memproduksi kalimat sendiri, membantu memahami teks, dan pemanfaatan kata berulang-ulang akan membuat kosakata tersebut familiar bagi pemelajar.

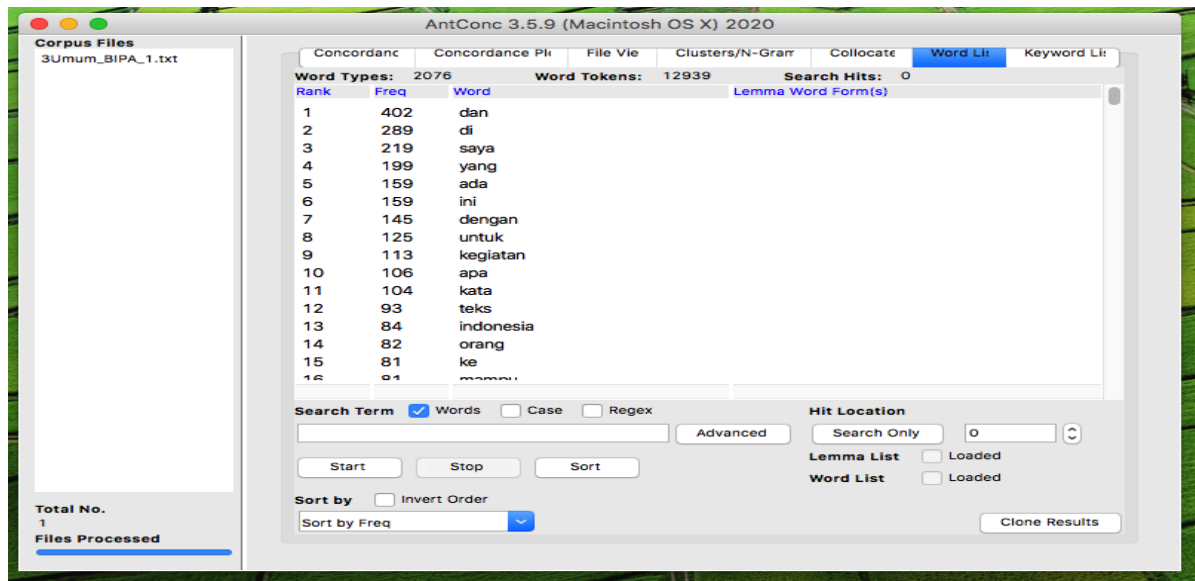
Kedua penelitian sebelumnya sama-sama memanfaatkan linguistik korpus untuk menentukan kata berfrekuensi tinggi. Perbedaannya terletak pada sumber/bahan korpus yang digunakan. Arum & Winarti (2020) menggunakan buku teks radiologi yang berjudul *Radiographic Positioning and Related Anatomy* karya Bontrager dan Lampignano sebagai korpus data dan Siagian, (2020) mendapatkan sumber data dari 17 buku BIPA tingkat pemula dari lembaga penyelenggara BIPA dari dalam dan luar negeri. Penelitian yang dilakukan peneliti bersumber dari buku BIPA “Sahabatku Indonesia BIPA 1” yang diterbitkan oleh BIPA Kemdikbud pada 2019. Pemilihan sumber data ini dikarenakan penggunaan bahan ajar ini dalam program BIPA Kemdikbud yaitu penugasan pengajar BIPA untuk luar negeri baik secara daring maupun luring.

Metode

Menurut Sudaryanto (2015), sebagai bagian dari serangkaian kegiatan ilmiah, kegiatan penelitian bahasa terbagi dalam dua kurun: pencarian masalah dan pemecahan masalah. Pencarian masalah diwujudkan dalam rumusan masalah. Kemudian masalah-masalah tersebut dipecahkan melalui tiga tahap, yaitu penyediaan data, penganalisisan data, dan penyajian hasil analisis data.

Data empiris dalam tahap penyediaan data diperoleh dari buku Sahabatku Indonesia BIPA 1 terbitan tahun 2019. Buku elektronik tersebut dapat diunduh secara gratis di laman bipa.kemdikbud.go.id (https://bipa.kemdikbud.go.id/belajar_formunduh.php?id=846) dengan mengisi formulir unduh yang tersedia. Pengambilan data dan penganalisisannya dimulai sejak 1 September 2021 serta berakhir pada 30 September 2021. Data berupa buku elektronik yang berjenis pdf harus diubah menjadi txt dengan menggunakan aplikasi pengubah fail. Dalam penelitian ini, peneliti mengunduh aplikasi “PDF to Text” dari *Apple Store* untuk mengubah fail pdf menjadi txt. Buku Sahabatku Indonesia BIPA 1 (fail pdf) diunggah ke dalam aplikasi tersebut dan berubah menjadi fail txt. Pengubahan ini dilakukan untuk seluruh bagian buku tidak hanya bagian materi/topik/unit saja.

Aplikasi *AntConc* yang digunakan dalam penelitian ini merupakan *AntConc 3.5.9* (Macintosh OS X) 2020. Aplikasi tersebut didapat secara gratis dengan cara mengunduh di laman <https://www.laurenceanthony.net/software/antconc/>. Setelah aplikasi *AntConc* terpasang, fail buku yang berbentuk txt diunggah ke dalam aplikasi *AntConc* (*tool-based corpus*). Untuk menentukan frekuensi kata, maka fitur yang digunakan dalam aplikasi *AntConc* adalah *Word List*. Dalam tahap pemerolehan data ini menggunakan metode kuantitatif melalui aplikasi *AntConc* tersebut. Berikut gambar aplikasi *AntConc* yang sudah memuat fail buku Sahabatku Indonesia BIPA 1 yang berbentuk txt.



Gambar 1. Aplikasi AntConc yang memuat fail (*txt*) buku Sahabatku Indonesia BIPA 1 (sumber dokumentasi pribadi)

Berdasarkan fitur *Word List*, terdapat 2.076 *word types* dengan 12.939 *word tokens*. Crawford & Eniko (2016) menyebutkan bahwa *word tokens are each word in a text, while word types are each type of word in a text* ‘token kata adalah setiap kata dalam teks, sedangkan jenis kata adalah setiap jenis kata dalam teks’. Berikut contoh kalimat untuk lebih memahami perbedaan *word token* dan *word type*. “Saya cinta Anda dan Anda cinta saya”, kalimat tersebut terdiri atas empat *word types* (saya, cinta, Anda, dan) dan tujuh *word tokens* (saya, cinta, Anda, dan, Anda, cinta, saya).

Selanjutnya, menganalisis data secara kualitatif, langkah pertama adalah mereduksi data. Data awal (2.076 *word types*) dianalisis berdasarkan kaidah bahasa Indonesia secara manual dan menyisakan 1.623 *word types*. Hasil tersebut didapat dengan menyisihkan data berupa:

1. nama diri;
2. nama tempat (negara/pulau/daerah);
3. akronim;
4. kata asing;
5. kata daerah (seperti *Teteh, Uda, Uni, Bli, Mbok*, dsb.); dan
6. kata yang tidak ditemukan artinya dalam KBBI.

Tahap kedua dalam menganalisis data adalah menentukan kelas kata berdasarkan KBBI V Daring secara manual. Hasil dari penganalisisan tersebut adalah ditemukannya kelas kata untuk 1.623 *word types*. Tahap penyajian data dengan menerapkan metode padan ortografis dengan alat penentu tulisan. Hasil analisis data pada penelusuran ini diuraikan dengan metode formal dan informal. Metode formal karena terdapat lambang-lambang sedangkan informal dengan penggunaan kata-kata biasa termasuk penggunaan terminologi yang bersifat teknis.

Hasil dan Pembahasan

Kata Berfrekuensi Tinggi pada Buku Sahabatku Indonesia BIPA 1

Berdasarkan analisis data menggunakan fitur *Word List* ‘daftar kata’ dalam AntConc, terdapat 2.076 *word types*. Setelah mereduksi data (sebanyak 2.076 *word types*), terdapat 1.623 jenis kata yang berbeda dan jumlah kemunculan kata tersebut mencapai 11.176. Hasil analisis data ini menampilkan

daftar kata yang diurutkan berdasarkan frekuensi penggunaannya dalam Buku Sahabatku Indonesia BIPA 1. Berikut sepuluh kata berfrekuensi tinggi pada buku tersebut yang dijelaskan dalam tabel 1.

Tabel 1. Sepuluh Kata Berfrekuensi Tinggi pada Buku Sahabatku Indonesia BIPA 1

Rank	Freq	Word
1	402	dan
2	289	di
3	219	Saya
4	199	Yang
5	159	Ada
6	159	Ini
7	145	Dengan
8	125	untuk
9	113	kegiatan
10	106	apa

Kolom I menunjukkan *rank* ‘peringkat’ kata berdasarkan frekuensi kata dalam data yang ditampilkan secara urut. Semakin kecil angka pada *rank* maka frekuensi kata semakin tinggi atau tingkat pemakaian/penggunaan/kemunculan kata yang tinggi/sering muncul. Kolom II menunjukkan *freq* ‘frekuensi’ yang merupakan jumlah pemakaian kata dalam data. Angka pada *freq* menunjukkan jumlah kemunculan kata dalam data. Kata *dan* muncul sebanyak 402 kali dalam buku Sahabatku Indonesia BIPA 1. Semakin ke bawah atau mendapatkan peringkat yang besar maka tingkat kemunculan kata juga semakin kecil. Kata *apa* menduduki peringkat ke-10 dengan tingkat kemunculan 106 kali.

Kata berfrekuensi tinggi yang ditemukan dalam buku Sahabatku Indonesia BIPA 1 yang diterbitkan pada 2019 berbeda dengan hasil penelitian kata berfrekuensi tinggi pada pembelajaran BIPA Pemula tahun 2020. Siagian (2020) menyebutkan bahwa jumlah kosakata yang ditemukan sebanyak 3.695. Frekuensi kata dengan tingkat kemunculan yang tinggi juga berbeda ketika dibandingkan dengan *Indonesian Leipzig Corpora Collection* tahun 2013. Sebagai *web-based corpus*, *Indonesian Leipzig Corpora Collection*, merupakan korpus data dalam bahasa Indonesia dari berbagai genre dengan jumlah *word types* 7.964.109 dan *word tokens* 1.206.281.985. Berikut tabel perbandingan sepuluh kata berfrekuensi tinggi dari ketiga sumber yang berbeda.

Tabel 2. Perbandingan Sepuluh Kata Berfrekuensi Tinggi

I	II	III
Indonesian Leipzig Corpora Collection (2013)	HFW Pembelajaran BIPA Pemula (Siagian: 2020)	Buku Sahabatku Indonesia BIPA 1 (Terbitan 2019)
1. yang	1. saya	1. dan
2. dan	2. di	2. di
3. di	3. ya	3. saya
4. dengan	4. dan	4. yang
5. ini	5. Anda	5. ini

6. untuk	6. ini	6. ada
7. dari	7. apa	7. dengan
8. itu	8. ada	8. untuk
9. dalam	9. ke	9. kegiatan
10. tidak	10. selamat	10. apa

Perbandingan tersebut menunjukkan bahwa ada tiga kata yang merupakan kata berfrekuensi tinggi dari ketiga sumber korpus yaitu *dan*, *di*, dan *ini*. Ketiga kata tersebut hanya memiliki perbedaan *rank* ‘peringkat’. Kolom II dan III memiliki persamaan yaitu sumber berasal dari buku BIPA pemula meskipun kolom II jenjang BIPA 1 dan 2 sedangkan kolom III jenjang BIPA 1 saja. Kata berfrekuensi tinggi yang sama tetapi berbeda peringkat dalam kolom II dan III meliputi: *saya*, *apa*, dan *ada*. Kolom III dan I juga terdapat kesamaan kata berfrekuensi tinggi yaitu kata *yang*, *dengan*, dan *untuk*. Kesepuluh kata berfrekuensi tinggi pada kolom I dan II berbentuk kata dasar. Pada kolom III terdapat satu kata berfrekuensi tinggi yang berbentuk kata turunan atau berimbuhan yaitu *kegiatan*. Meskipun berbeda sumber data atau korpus yang digunakan namun persamaan terletak pada bahasa yaitu bahasa Indonesia.

Berikut tabel 3 dijelaskan mengenai daftar 100 kata berfrekuensi tinggi yang sebaiknya diperkenalkan secara bertahap kepada pemelajar BIPA 1 sehingga pemelajar lebih semangat dalam memahami teks yang terdapat dalam buku Sahabatku Indonesia BIPA 1.

Tabel 3. Seratus Kata Berfrekuensi Tinggi dalam Buku Sahabatku Indonesia BIPA 1

1	402	dan	37	45	Dia	73	26	jika
2	289	di	38	44	Teman	74	26	lokasi
3	219	saya	39	43	Ulang	75	26	pengembangan
4	199	yang	40	42	aktivitas	76	26	berwarna
5	159	ini	41	42	Tempat	77	25	ciri
6	159	ada	42	41	Contoh	78	25	secara
7	145	dengan	43	40	pukul	79	24	Berbicara
8	125	untuk	44	40	kami	80	23	Baik
9	113	kegiatan	45	39	tentang	81	23	Lalu
10	106	apa	46	38	akan	82	23	Setelah
11	104	kata	47	37	pendek	83	23	Deskripsi
12	93	teks	48	37	membaca	84	23	Dua
13	82	orang	49	36	harian	85	23	Kebutuhan
14	81	mampu	50	36	pertanyaan	86	23	Bagaimana
15	81	ke	51	35	atau	87	23	Pergi
16	80	anda	52	35	informasi	88	22	Benar
17	78	dalam	53	35	mana	89	22	Materi
18	76	pada	54	35	menulis	90	22	Tanggal
19	75	sumber	55	34	sifat	91	21	Konkret
20	69	sederhana	56	31	keluarga	92	21	Saja
21	66	bahasa	57	30	hobi	93	21	Luh
22	66	menggunakan	58	30	makan	94	21	Ni
23	63	tahun	59	30	tidur	95	21	Badan
24	62	dari	60	30	tulis	96	21	Depan
25	60	audio	61	29	bulan	97	21	Kamu

26	59	hari	62	28	sangat	98	21	Lahir
27	57	rumah	63	28	jawab	99	21	Menanyakan
28	56	itu	64	28	waktu	100	21	Menyapa
29	55	berkaitan	65	28	bisa			
30	51	ungkapan	66	28	simak			
31	49	tidak	67	27	lain			
32	49	kalimat	68	27	arah			
33	49	unit	69	27	nama			
34	47	bahan	70	27	tujuan			
35	46	suka	71	27	aku			
36	45	juga	72	27	adalah			

Frekuensi Kemunculan Kelas Kata pada Buku Sahabatku Indonesia BIPA 1

Frekuensi kemunculan kelas kata diperoleh dengan menganalisis kelas kata dari 1.623 data secara manual menggunakan aplikasi KBBI V Daring. Analisis kelas kata berdasarkan kelas kata dalam KBBI V Daring. Daftar kelas kata dalam KBBI V Daring berdasarkan petunjuk pemakaian KBBI edisi kelima, yang meliputi:

- (1) a adjektiva, yaitu kata yang menjelaskan nomina atau pronomina;
- (2) adv adverbia, yaitu kata yang menjelaskan verba, adjektiva, adverbia lain, atau kalimat;
- (3) n nomina, yaitu kata benda;
- (4) num numeralia, yaitu kata bilangan;
- (5) p partikel, kelas kata yang meliputi kata depan, kata sambung, kata seru, kata sandang, ucapan salam;
- (6) pron pronomina, kelas kata yang meliputi kata ganti, kata tunjuk, dan kata tanya; dan
- (7) v verba, yaitu kata kerja

Berikut gambar 2 yang menunjukkan contoh analisis kelas kata menggunakan KBBI V Daring dengan memasukkan daftar kata yang didapat dari pengolahan data melalui fitur *Word List AntConc*. Kata *saya* ketika dimasukkan ke dalam KBBI V Daring akan menghasilkan keterangan *pron* sebelum definisi kata. *Pron* berarti kata *saya* merupakan bagian dari kelas kata *pronomina*.

The screenshot shows a web browser window with the URL kbbi.kemdikbud.go.id/entri/saya. The page header includes navigation links like 'Gmail', 'Maps', 'Translate', 'Pencarian - KBBI...', 'Bahasa Indonesia', 'Pengajaran Bahasa', and 'Ebook'. Below the header, there is a search bar with the text 'saya' and a magnifying glass icon. The search results display 'sa.ya' with a sub-link '→ Tesaurus'. Underneath, there is an 'Etimologi:' section with a note in brackets: '[Informasi etimologi hanya tersedia bagi pengguna terdaftar]'. The etymology is listed as: 1. *pron* orang yang berbicara atau menulis (dalam ragam resmi atau biasa); aku; 2. *pron* ya: --, Tuan!

Gambar 2. Tahap penentuan kelas kata dari data menggunakan KBBI V Daring (sumber dokumentasi pribadi)

Berikut tabel 4 dijelaskan mengenai hasil analisis kelas kata dan dilengkapi dengan jumlah kata (*word types*) serta jumlah kemunculan (*word tokens*) pada Buku Sahabatku Indonesia BIPA 1.

Tabel 4. Jumlah Kata dan Kemunculan Kelas/Kategori Kata pada Buku Sahabatku Indonesia BIPA 1

Kelas/Kategori Kata	Word Types	Word Token
Nomina	855	4.783
Verba	426	2.134
Partikel	52	1.756
Adjektiva	174	948
Pronomina	28	923
Adverbia	52	415
Numeralia	33	207
Klitika	3	10
Total	1.623	11.176

Berdasarkan analisis, kelas kata dengan frekuensi tertinggi ditempati oleh *nomina* dengan total kemunculan 4.783. Kemudian, ada kelas kata *verba* dengan frekuensi kemunculan sebesar 2.134. Kelas kata berikutnya yang memiliki kemunculan yang tinggi adalah *partikel* yaitu 1.756. *Partikel* yang memiliki frekuensi tinggi ternyata hanya terdapat 52 *word types* dalam buku Sahabatku Indonesia BIPA 1. Dalam analisis kelas kata berfrekuensi tinggi ditemukan kelas kata *klitik* dalam jumlah yang tidak signifikan. Oleh karena itu, terdapat delapan kelas kata berdasarkan analisis *Word List* ‘daftar kata’ pada buku Sahabatku Indonesia BIPA 1.

Berikut pada tabel 4 dijelaskan urutan kelas kata berdasarkan frekuensi kemunculannya (tinggi ke rendah) dan dilengkapi dengan lima kata berfrekuensi tertinggi di masing-masing kelas kata tersebut.

Tabel 4. Kosakata Berfrekuensi Tinggi Berdasarkan Kelas Kata

Kelas Kata	Kata
Nomina (n)	kegiatan (113); kata (104); teks (93); orang (82); sumber (75)
Verba (v)	ada (159); menggunakan (66); berkaitan (55); suka (46); ulang (43)
Partikel (p)	dan (402); di (289); yang (199); dengan (145); untuk (125)
Adjektiva (a)	mampu (81); sederhana (69); pendek (37); lain (27); baik (23)
Pronomina (pron)	saya (219); ini (159); apa (106); Anda (80); itu (56)
Adverbia (adv)	tidak (49); juga (45); akan (38); sangat (28); saja (21)
Numeralia (num)	dua (23); empat (19); lima (18); tiga (18); beberapa (16)
Klitik	ku (4); mu (4); nya (2)

Berdasarkan tabel 4, kata berfrekuensi tinggi dalam buku Sahabatku Indonesia BIPA 1 yaitu *dan*, *di*, *yang*, *dengan*, *untuk* merupakan *partikel*. Penyumbang kata berfrekuensi tinggi berikutnya adalah dari kelas kata *pronomina* yaitu kata *saya*, *ini*, dan *apa*. Kelas kata *nomina* dengan kata *kegiatan* serta kelas kata *verba* dengan kata *ada*. Dengan demikian, pengenalan kata berfrekuensi tinggi terutama

dari kelas kata partikel sebaiknya disampaikan kepada pemelajar BIPA 1 guna menunjang mereka dalam memahami bahan ajar secara mandiri atau melalui tutorial/bimbingan.

Simpulan

Hasil penelitian menyuratkan jika kelas kata yang paling tinggi tingkat kemunculannya tidak serta merta menyumbang kata berfrekuensi tinggi. Dalam penelitian ini, *partikel* merupakan penyumbang terbanyak dari kata berfrekuensi tinggi (*dan, di, yang, dengan, untuk*). Maka, dapat disimpulkan bahwa daftar kata berfrekuensi tinggi terkadang kata yang tidak merupakan kata teknik berdasarkan topik yang diajarkan. Penelitian menentukan kata berfrekuensi tinggi dengan memanfaatkan linguistik korpus bisa diterapkan ke berbagai bahan (buku, koran, bacaan, dsb.) bergantung kebutuhan. Penelitian ini dapat menghasilkan daftar kata khusus ataupun istilah yang dikemas dalam bentuk kamus mini yang bisa dijadikan bahan rujukan oleh pemelajar. Selain itu, penelitian ini mendeskripsikan kata-kata yang didukung oleh fakta berupa angka (frekuensi kemunculan). Serta beberapa data yang direduksi karena salah penulisan dapat dijadikan sebagai koreksi untuk penulis bahan ajar.

Ucapan Terima Kasih

Peneliti mengucapkan terima kasih khususnya kepada dosen pembimbing, Bapak Dr. Tri Mastoyo Jati Kesuma, M.Hum., peneliti yang telah memberikan saran dalam penulisan artikel ini. Selain itu, peneliti mengucapkan terima kasih kepada BIPA Kemdikbud yang telah menyiapkan bahan ajar BIPA berbagai kategori untuk diunduh secara gratis.

Daftar Rujukan

- Anthony, L. (2004). AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit. *Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning*, 7–13.
- Arum, E. R., & Winarti, W. (2020). Penggunaan Linguistik Korpus dalam Mempersiapkan Bahan Ajar English For Specific Purpose Di Bidang Radiologi. *Jurnal Teras Kesehatan*, 2(2), 58–69. <https://doi.org/10.38215/jutek.v2i2.39>
- Crawford, W. J., & Eniko, C. (2016). *Doing Corpus Linguistics*. Routledge.
- Hizbullah, N., Fazlurrahman, & Fuzi Fauziah. (2016). Linguistik Korpus dalam Kajian dan Pembelajaran bahasa Arab di Indonesia. *Prosiding Konferensi Nasional Bahasa Arab II*, 385–393.
- Johns, J. L., & Wilke, K. H. (2018). High Frequency Words: Some Ways to Teach and Help Students Practice and Learn Them. *Texas Journal of Literacy Education* |, 6(1), 3–13. <https://www.mendeley.com/viewer/?fileId=354e4718-bc11-5edb-42f7-a39bc4db1f3a&documentId=136393f0-fbe8-3298-aefa-ff72f5938f0d>
- Lightbown, P. M., & Spada, N. (2011). *How Languages are Learned Third Edition*. University Press.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge University Press.
- Siagian, E. N. (2020). Kata Berfrekuensi Tinggi dalam Pembelajaran BIPA Pemula. *Ranah: Jurnal Kajian Bahasa*, 9(2), 188–201.
- Sudaryanto. (2015). *Metode dan Aneka Teknik Analisis Bahasa: Pengantar Penelitian Wahana Kebudayaan secara Linguistik*. Sanata Dharma University Press.