Kata Berfrekuensi Tinggi
Dalam Korpus Pemelajar Bahasa Indonesia Beranotasi Eror

エラータグ付きインドネシア語学習者コーパスにおける高頻度語

David Moeljadi

ダヴィド・ムルヤディ

Abstract

This study aims to give a list of high frequency words in Indonesian used by Indonesian language learners at all levels in writing essays at seven universities in Japan. This study uses Koper (Korpus Pemelajar Bahasa Indonesia Beranotasi Error), the Error-Annotated Indonesian Language Learners Corpus. Koper contains a collection of essays written by students taking Indonesian language courses at seven universities in Japan: Kanda University of International Studies (KUIS), Ritsumeikan Asia Pacific University (APU), Tokyo University of Foreign Studies (TUFS), Osaka University, Chuo University, Keio University, and Sophia University. The essays used in this study were collected from April 2023 to March 2024. The result of this study will be useful for Indonesian language learners and teachers. The high frequency words in this corpus can be used in Indonesian language teaching, for example in determining the list of vocabulary that must be taught or targeted according to the level.

1. Pendahuluan

Kosakata memiliki peran sangat penting dalam pemelajaran bahasa asing, meliputi semua aspek keterampilan berbahasa, yaitu keterampilan membaca, mendengarkan, menulis, dan berbicara. Semakin banyak kosakata yang dikuasai oleh pemelajar, semakin tinggi tingkat kemampuan berbahasa asing pemelajar tersebut. Siagian (2020) menulis bahwa pemelajaran kosakata perlu mempertimbangkan konteks, kata-kata berfrekuensi tinggi (*high frequency words*), kata-kata akademis (*academic words*), dan kata-kata teknis (*technical words*) yang bergantung pada level dan tujuan pemelajaran. Kata berfrekuensi tinggi terutama menentukan keberhasilan pemelajar dalam penguasaan bahasa asing. Kata-kata yang lebih sering muncul dalam teks cenderung lebih mudah diingat oleh pemelajar. Berdasarkan penelitian Sanchez Gutierrez dkk. (2019), kata yang muncul sepuluh kali lebih mudah diingat dibandingkan kata yang hanya muncul lima kali. Selain itu, pengulangan dalam pemelajaran membantu memperkuat daya ingat pemelajar serta meningkatkan motivasi dan kepercayaan diri pemelajar dalam memproduksi kalimat (Siagian, 2020). Kamus kata berfrekuensi tinggi bahasa Prancis, Jepang, dan Mandarin berperan penting dalam pemelajaran bahasa asing tersebut (Purwiyanti dkk., 2017).

Dalam konteks pembelajaran Bahasa Indonesia bagi Penutur Asing (BIPA), penelitian tentang kata berfrekuensi tinggi dalam bahasa Indonesia belum banyak dilakukan, apalagi penyusunan kamus kata berfrekuensi tinggi BIPA. Muhadjir dkk. (1996) meneliti frekuensi kosakata bahasa Indonesia. Penelitian tersebut menghasilkan 3.800 kosakata dasar yang frekuensi penggunaannya 20 kali atau lebih dengan sumber data dari surat kabar Kompas tahun 1994, diambil secara acak sekali dalam seminggu. Hananto dkk. (2011) melakukan penelitian *Frequency-Based Indonesian Word List* dengan sumber data buku-buku

SMP, cerita fiksi anak-anak dan dewasa, serta artikel koran. Selain itu, daftar 50 kata berfrekuensi tinggi dalam bahasa Indonesia dapat diakses dari situs Corpora and Language Statistics (Eckart dan Quasthoff, 2013) dengan sumber data korpus Leipzig (Leipzig Corpora Collection) yang diambil dari artikel Wikipedia dan artikel daring bahasa Indonesia lainnya. Siagian (2020) meneliti kata berfrekuensi tinggi dalam BIPA pemula dengan sumber data materi percakapan dari 17 buku BIPA 1 dan 2 tingkat pemula. Penelitian tersebut menghasilkan 3.695 kata berfrekuensi tinggi.

Moeljadi (2023) mengembangkan korpus pemelajar bahasa Indonesia pertama yang bernama Koper (Korpus Pemelajar Bahasa Indonesia Beranotasi Eror). Koper adalah koleksi tulisan pemelajar bahasa Indonesia yang dapat dibaca komputer dan berisi data pemelajar. Tulisan pemelajar bahasa Indonesia ini dikumpulkan sejak April 2023 dari tujuh universitas di Jepang. Keberadaan Koper ini sangat penting untuk memahami kecenderungan penggunaan kosakata bahasa Indonesia oleh pemelajar BIPA, terutama di Jepang. Penelitian ini berfokus pada kata berfrekuensi tinggi dalam Koper. Dengan menggunakan Koper, dapat dianalisis kecenderungan penggunaan kosakata dalam tulisan pemelajar BIPA. Hasil penelitian ini diharapkan dapat memberi manfaat bagi pemelajar, pengajar, penulis bahan ajar, dan pihak-pihak yang terkait dengan BIPA.

2. Sumber Data dan Pengolahan Data

Koper berisi tulisan pemelajar BIPA dari berbagai level yang dikumpulkan dari tujuh universitas di Jepang, yaitu Kanda University of International Studies, Ritsumeikan Asia Pacific University, Tokyo University of Foreign Studies, Chuo University, Keio University Shonan Fujisawa Campus, Sophia University, dan Osaka University.[1] Penulis meminta bantuan dosen bahasa Indonesia di tiap universitas tersebut untuk mengumpulkan tulisan atau karangan pemelajar. Tulisan pemelajar dikumpulkan baik dalam bentuk tulisan tangan maupun hasil ketikan. Selain karangan atau tulisan pemelajar, data pemelajar yang meliputi nama, umur, jenis kelamin, lama belajar bahasa Indonesia, pengalaman tinggal di Indonesia, dan sebagainya dikumpulkan beserta dengan formulir persetujuan. Tidak semua tulisan dari semua pemelajar dimasukkan sebagai data dalam Koper. Tulisan yang berhasil dikumpulkan adalah tulisan pemelajar yang bersedia menyumbangkan tulisannya.

Semua tulisan yang terkumpul diseragamkan ke dalam format teks (.txt). Informasi pribadi yang ada dalam tulisan pemelajar seperti nama, nomor telepon, dan alamat pos-el diganti dengan label [NAMA], [NOMOR], [POS-EL]. Tulisan atau karangan dikelompokkan menurut sumber data: kode nama penulis, nama universitas, level atau tingkat pemelajar, tema tulisan/karangan, dan tahun penulisan. Level atau tingkat pemelajar dibagi ke dalam empat level berdasarkan tingkat kuliah mahasiswa S-1 jurusan atau program studi bahasa Indonesia di KUIS, TUFS, dan OU. Pembagian level pemelajar di universitas yang tidak memiliki jurusan atau program studi bahasa Indonesia tetapi memiliki mata kuliah pilihan bahasa Indonesia, yaitu APU, CHU, SFC, dan SOP, adalah menurut level mata kuliah di tiap universitas tersebut. APU memiliki empat level kelas bahasa Indonesia yang bisa disamakan dengan empat tingkat kuliah.

[1] Dalam makalah ini, Kanda University of International Studies disingkat menjadi KUIS, Ritsumeikan Asia Pacific University disingkat menjadi APU, Tokyo University of Foreign Studies disingkat menjadi TUFS, Chuo University disingkat menjadi CHU, Keio University Shonan Fujisawa Campus disingkat menjadi SFC, Sophia University disingkat menjadi SOP, dan Osaka University disingkat menjadi OU.

CHU hanya memiliki satu level, yaitu level dasar (*shokyu*) yang bisa disamakan dengan tingkat satu. SOP memiliki dua level, yaitu level dasar (*shokyu*) dan level menengah (*chukyu*) yang bisa disamakan dengan tingkat satu dan tingkat dua. SFC memiliki lima kelas bahasa Indonesia, yaitu Basic 1 (*kisokyu*), Basic 2 (*shokyu*), Intensif 1 (*kiso shokyu*), Intensif 2 (*shochukyu*), dan Skill (*chukyu*). Basic 1, Basic 2, dan Intensif 1 dapat disamakan dengan tingkat satu, Intensif 2 dapat disamakan dengan tingkat dua, dan Skill dapat disamakan dengan tingkat tiga.

Penulis menggunakan program pemisah kata dan penghitung frekuensi, alih-alih perangkat lunak AntConc (Anthony, 2024) yang biasanya digunakan untuk memproses dan menganalisis korpus. AntConc mengelompokkan kata berdasarkan spasi dan tanda hubung. Perangkat lunak ini tidak mengenal kata majemuk, jadi kata majemuk seperti ‘rumah sakit’ dan ‘air mata’ akan dipisah menjadi ‘rumah’, ‘sakit’, ‘air’, dan ‘mata’ berdasarkan penanda spasi. Selain itu, AntConc tidak mengenal kata ulang, jadi kata ulang seperti ‘laki-laki’ akan dipisah menjadi ‘laki’ dan ‘laki’ tanpa tanda hubung. Berbeda dengan AntConc, program pemisah kata dan penghitung frekuensi yang digunakan dalam penelitian ini mengelompokkan kata berdasarkan penanda spasi dan karakter atau tanda baca lainnya seperti tanda kurung buka, tanda titik dua, dan tanda tanya. Huruf kapital di awal kalimat dinormalisasi menjadi huruf kecil. Program tersebut juga menggunakan data kata majemuk dari KBBI sehingga kata-kata majemuk dapat dikenali dan unsur-unsur pembentuknya tidak dipisah. Selain itu, unsur pembentuk kata ulang juga tidak dipisah. Kata-kata yang berhasil diekstrak kemudian diurutkan berdasarkan kemunculannya di dalam korpus. Setelah itu, kata-kata berfrekuensi tinggi yang berhasil diperoleh tersebut dikelompokkan berdasarkan kelas kata.

3. Hasil Penelitian dan Analisis

Penulis berhasil mengumpulkan 1.427 tulisan yang berisi 177.705 kata dari 351 mahasiswa yang merupakan pemelajar BIPA di 7 universitas di Jepang selama April 2023 hingga Maret 2024.[2] Tabel 1 menunjukkan rincian jumlah tulisan, kata, dan pemelajar per tingkat.

Tabel 1. Jumlah Tulisan Pemelajar BIPA yang Terkumpul April 2023—Maret 2024

Tingkat	Teks		Kata		Pemelajar	
	Jumlah	Persentase	Jumlah	Persentase	Jumlah	Persentase
1	600	42%	36.551	21%	216	62%
2	608	43%	101.575	57%	95	27%
3	180	13%	33.008	19%	33	9%
4	39	3%	6.571	4%	7	2%
Total	1.427	100%	177.705	100%	351	100%

Jika dilihat dari jumlah pemelajar, dapat dilihat bahwa pemelajar tingkat satu yang terbanyak, yaitu sebanyak 62%. Namun, jika dilihat dari jumlah tulisan dan kata, dapat dilihat bahwa jumlah tulisan pemelajar tingkat dua yang terbanyak dengan jumlah kata (57% dari keseluruhan) hampir tiga kali lebih

[2] Tahun perkuliahan di Jepang dimulai pada bulan April dan selesai pada bulan Maret.

banyak daripada tulisan pemelajar tingkat satu (21%) dan tingkat tiga (19%).

Tabel 2. Jumlah Tulisan Pemelajar BIPA di KUIS, OU, dan TUFS April 2023—Maret 2024

	KUIS			OU			TUFS		
Tingkat	Teks	Kata	Pemelajar	Teks	Kata	Pemelajar	Teks	Kata	Pemelajar
1	49	5.405	17	0	0	0	0	0	0
2	26	1.673	25	32	13.096	16	399	72.782	22
3	32	5.678	5	0	0	0	120	23.991	16
4	35	6.225	5	0	0	0	0	0	0
Total	142	18.981	52	32	13.096	16	519	96.773	38

Tabel 2 menunjukkan rincian jumlah tulisan, kata, dan pemelajar per tingkat di KUIS, OU, dan TUFS. Ketiga universitas ini memiliki jurusan atau program studi bahasa Indonesia tingkat S-1. Mahasiswa atau pemelajar BIPA di universitas tersebut belajar bahasa Indonesia selama empat tahun. Penulis berhasil mengumpulkan tulisan pemelajar di KUIS dari tingkat satu hingga empat. Tema tulisan pemelajar tiap tingkat berbeda-beda. Pemelajar tingkat satu menulis karangan tentang perkenalan diri, keluarga, rumah, dan cerita lucu. Pemelajar tingkat dua menulis surat cinta. Pemelajar tingkat tiga dan empat menulis karangan tentang tokoh, tempat, perjalanan, pos-el, paragraf perbandingan, opini, dan pendahuluan makalah. Dari pemelajar di OU, penulis hanya dapat mengumpulkan tulisan pemelajar tingkat dua yang menulis tentang tokoh dan tempat wisata. Dari pemelajar di TUFS, penulis berhasil mengumpulkan tulisan pemelajar tingkat dua dan tiga. Pemelajar tingkat dua menulis tentang tokoh, teman, tempat, budaya, cerita rakyat, hobi, keluarga, liburan, pidato, dan rumah. Pemelajar tingkat tiga menulis naskah pidato, surat formal, undangan, tokoh, tempat, perjalanan, hobi, cerita rakyat, dan pengalaman.

Tabel 3. Jumlah Tulisan Pemelajar BIPA di APU, CHU, SFC, dan SOP April 2023—Maret 2024

	APU			CHU			SFC			SOP		
Tingkat	Teks	Kata	Pljar	Teks	Kata	Pljar	Teks	Kata	Pljar	Teks	Kata	Pljar
1	91	6.238	46	157	9.874	51	205	8.500	83	98	6.534	19
2	12	1.843	5	0	0	0	75	4.700	15	64	7.481	12
3	14	1.924	5	0	0	0	14	1.415	7	0	0	0
4	4	346	2	0	0	0	0	0	0	0	0	0
Total	121	10.351	58	157	9.874	51	294	14.615	105	162	14.015	31

Tabel 3 menunjukkan rincian jumlah tulisan, kata, dan pemelajar per tingkat di APU, CHU, SFC, dan SOP. Keempat universitas ini memiliki kelas bahasa Indonesia sebagai mata kuliah pilihan bagi mahasiswa S-1. Penulis berhasil mengumpulkan tulisan pemelajar di APU dari level satu hingga level empat. Pemelajar tingkat satu menulis tentang keluarga, teman, tempat, makanan, dan film. Pemelajar tingkat dua menulis tentang teman, tempat, dan keluarga. Pemelajar tingkat tiga menulis tentang kesan wawancara, keluarga, dan rumah. Pemelajar tingkat empat menulis pos-el formal dan karangan tentang

gempa dan kegiatan selama liburan. Dari pemelajar BIPA di CHU, penulis mengumpulkan tulisan pemelajar kelas dasar (*shokyu*) bertemakan keluarga, teman, tempat, tokoh, makanan, dan Indonesia.

Dari pemelajar BIPA di SFC, penulis mengumpulkan tulisan pemelajar tingkat satu tentang perkenalan diri, rumah, teman, dan hobi; tulisan pemelajar tingkat dua tentang perjalanan, tradisi, dan makanan; serta tulisan pemelajar tingkat tiga tentang liburan, tempat, dan film. Penulis juga mengumpulkan tulisan pemelajar tingkat satu dan dua di SOP. Tulisan pemelajar tingkat satu tentang tokoh, tempat, keluarga, makanan, aktivitas, dan liburan. Tulisan pemelajar tingkat dua tentang perjalanan, rumah idaman, sahabat, dan masa kecil.[3] Setelah semua data tulisan pemelajar ini terkumpul, penulis menggunakan program pemisah kata dan penghitung frekuensi untuk memperoleh daftar kata berfrekuensi tinggi secara keseluruhan dan per tingkat pemelajar.

Tabel 4. Sepuluh Kata Berfrekuensi Tertinggi

	Muhadjir dkk. (1996)	Hananto dkk. (2011)	Eckart dan Quasthoff (2013)	Siagian (2020)
1	yang	yang	yang	saya
2	dan	dan	dan	di
3	di	di	di	ya
4	itu	dengan	dengan	dan
5	dengan	itu	ini	Anda
6	ini	ini	untuk	ini
7	dari	dari	dari	apa
8	dalam	tidak	itu	ada
9	untuk	untuk	dalam	ke
10	tidak	dalam	tidak	selamat

Tabel 4 berisi sepuluh kata berfrekuensi tertinggi yang diperoleh dari penelitian sebelumnya, yaitu Muhadjir dkk. (1996), Hananto dkk. (2011), Corpora and Language Statistics (Eckart dan Quasthoff, 2013) dengan sumber data korpus Leipzig (Leipzig Corpora Collection), dan Siagian (2020). Penelitian Muhadjir dkk. (1996), Hananto dkk. (2011), dan Eckart dan Quasthoff (2013) menggunakan korpus yang datanya berupa sumber tertulis beragam formal, yaitu artikel surat kabar Kompas, buku teks SMP, cerita anak, novel, artikel Wikipedia, dan artikel daring lainnya. Sebaliknya, penelitian Siagian (2020) menggunakan korpus yang datanya berupa materi percakapan buku teks BIPA pemula yang cenderung beragam informal.

Dari Tabel 4 di atas dapat dilihat bahwa tiga kata berfrekuensi tertinggi dalam bahasa Indonesia ragam tulis formal adalah pronomina relatif ‘yang’, konjungsi ‘dan’, dan preposisi ‘di’. Berikutnya, pronominal penunjuk (demonstrativa) ‘itu’, ‘ini’, dan preposisi ‘dengan’ menduduki peringkat keempat hingga keenam dari hasil penelitian Muhadjir dkk. (1996) dan Hananto dkk. (2011). Preposisi ‘dari’, ‘dalam’, ‘untuk’, dan kata negasi ‘tidak’ menduduki peringkat keenam hingga kesepuluh dari hasil penelitian Muhadjir dkk. (1996), Hananto dkk. (2011), dan Eckart dan Quasthoff (2013) walaupun urutannya

[3] Tulisan-tulisan pemelajar ini dapat diunduh di https://github.com/davidmoeljadi/Koper/tree/main

berbeda.

Sebaliknya, sepuluh kata berfrekuensi tertinggi bahasa Indonesia ragam percakapan (informal) dari hasil penelitian Siagian (2020) adalah sebagai berikut: pronomina persona ‘saya’, preposisi ‘di’, partikel ‘ya’, konjungsi ‘dan’, pronomina persona ‘Anda’, pronomina penunjuk (demonstrativa) ‘ini’, kata tanya ‘apa’, verba ‘ada’, preposisi ‘ke’, dan kata ‘selamat’. Hasil penelitian ini menunjukkan bahwa pronomina persona ‘saya’ dan ‘Anda’ banyak digunakan dalam percakapan. Dalam konteks BIPA pemula, memang pronomina persona ‘saya’ dan ‘Anda’ lebih banyak digunakan daripada ‘aku’ dan ‘kamu’. Selain itu, kata tanya ‘apa’, partikel ‘ya’ dalam jawaban, dan kata ‘selamat’ dalam ucapan salam sering digunakan dalam percakapan. Baik dalam ragam tulis formal maupun dalam ragam percakapan, preposisi ‘di’, konjungsi ‘dan’, dan pronomina penunjuk ‘ini’ menduduki peringkat atas kata berfrekuensi tinggi.

Tabel 5. Sepuluh Kata Berfrekuensi Tertinggi dalam Koper April 2023—Maret 2024

	Kata	Kelas Kata	Jumlah	Persentase
1	saya	pronomina	8.079	4,63%
2	di	preposisi	6.572	3,77%
3	dan	konjungsi	5.501	3,15%
4	dia	pronomina	3.667	2,10%
5	yang	pronomina	3.608	2,07%
6	itu	pronomina	2.471	1,42%
7	ke	preposisi	2.254	1,29%
8	ini	pronomina	1.956	1,12%
9	ada	verba	1.870	1,07%
10	dengan	preposisi	1.847	1,06%

Tabel 5 memuat sepuluh kata berfrekuensi tertinggi secara keseluruhan yang diperoleh penulis dari Koper dari April 2023 hingga Maret 2024.[4] Dapat dilihat bahwa ada kesamaan dengan hasil penelitian Siagian (2020), yaitu pronomina persona ‘saya’ dan preposisi ‘di’ menduduki peringkat pertama dan kedua. Selain itu, preposisi ‘ke’ dan verba ‘ada’ juga ada dalam daftar sepuluh kata berfrekuensi tertinggi. Selain itu, ada juga kesamaan dengan hasil penelitian Muhadjir dkk. (1996), Hananto dkk. (2011), dan Eckart dan Quasthoff (2013), yaitu penggunaan pronomina relatif ‘yang’, pronomina penunjuk ‘itu’, dan preposisi ‘dengan’. Hal ini menunjukkan bahwa tulisan pemelajar BIPA memiliki ciri khas ragam tulis formal dan berfokus pada diri sendiri yang dapat dilihat dari banyaknya penggunaan pronomina persona pertama tunggal ‘saya’. Hal ini tidak lepas dari tema atau topik tulisan pemelajar BIPA yang kebanyakan mengenai perkenalan diri, keluarga, hobi, kesan dan opini pribadi, dan sebagainya. Perbedaan dengan penelitian sebelumnya adalah pada pronomina persona ketiga tunggal ‘dia’ yang terutama digunakan ketika menulis tentang teman dan tokoh. Berbeda dengan ragam percakapan, pronomina persona kedua tunggal ‘Anda’ hanya digunakan dalam penulisan pos-el, undangan, surat formal, dan naskah pidato.

[4] Daftar selengkapnya dapat diunduh di https://github.com/davidmoeljadi/Koper/tree/main/HFW

Tabel 6. Kata Berfrekuensi Tinggi dalam Koper April 2023—Maret 2024 berdasarkan Tingkat Pemelajar

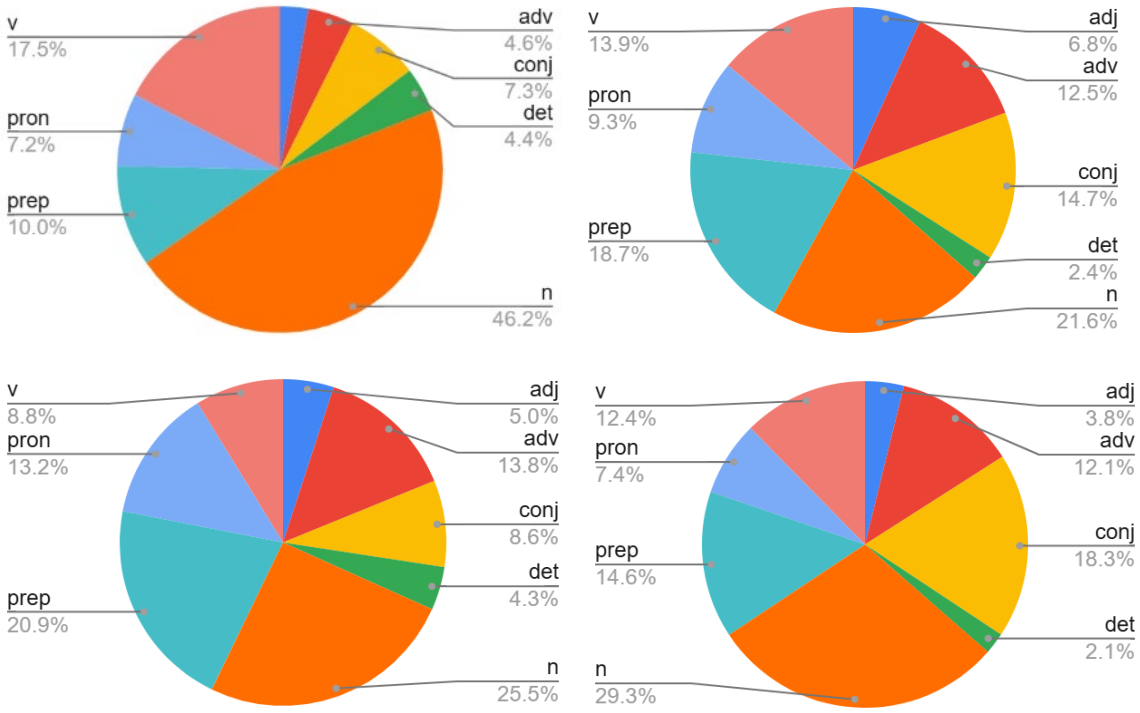
	Tingkat 1			Tingkat 2			Tingkat 3			Tingkat 4		
	Kata	Frek.	%	Kata	Frek.	%	Kata	Frek.	%	Kata	Frek.	%
1	saya	2.565	6,98	saya	3.930	3,96	saya	1.270	3,95	saya	314	5,01
2	di	1.716	4,67	di	3.547	3,58	dan	1.099	3,42	di	234	3,73
3	dia	1.599	4,35	dan	3.135	3,16	di	1.075	3,34	dan	188	3,00
4	dan	1.079	2,94	yang	2.561	2,58	yang	743	2,31	yang	105	1,67
5	jam	685	1,86	dia	1.807	1,82	itu	556	1,73	itu	101	1,61
6	ada	628	1,71	itu	1.655	1,67	ke	376	1,17	Indonesia	94	1,50
7	ke	505	1,37	ke	1.302	1,31	ada	370	1,15	dengan	91	1,45
8	dari	497	1,35	ini	1.166	1,18	ini	365	1,13	karena	83	1,32
9	[NAMA]	490	1,33	dengan	1.101	1,11	kami	362	1,13	tidak	73	1,16
10	nama	489	1,33	adalah	1.037	1,05	dengan	353	1,10	ke	71	1,13
11	tinggal	444	1,21	dari	830	0,84	untuk	329	1,02	Jepang	67	1,07
12	ini	370	1,01	ada	822	0,83	karena	280	0,87	adalah	57	0,91
13	makan	338	0,92	untuk	802	0,81	tidak	276	0,86	ini	55	0,88
14	pergi	336	0,91	banyak	782	0,79	adalah	274	0,85	untuk	55	0,88
15	ibu	306	0,83	Jepang	782	0,79	pada	260	0,81	orang	52	0,83
16	dengan	302	0,82	karena	776	0,78	sangat	251	0,78	ada	50	0,80

Tabel 6 berisi daftar kata berfrekuensi tinggi yang dikelompokkan menurut tingkat pemelajar. Kata yang paling banyak digunakan oleh pemelajar dari tingkat satu hingga empat adalah pronomina persona pertama tunggal ‘saya’. Tabel 6 menunjukkan perbedaan penggunaan kosakata dalam tulisan pemelajar tingkat satu dan pemelajar tingkat dua hingga empat. Pemelajar tingkat satu masih jarang menggunakan pronomina relatif ‘yang’. Sebaliknya, pemelajar tingkat dua hingga empat banyak menggunakan kata tersebut. Dapat dilihat dari Tabel 6 bahwa kata ‘yang’ menduduki peringkat keempat. Selain itu, kata ‘itu’ juga banyak digunakan oleh pemelajar tingkat dua hingga empat, tetapi jarang digunakan oleh pemelajar tingkat satu. Perbedaan lain yang dapat dilihat adalah penggunaan verba kopula ‘adalah’ dan konjungsi subordinatif ‘karena’. Verba ‘adalah’ dan konjungsi ‘karena’ mulai banyak digunakan dalam tulisan oleh pemelajar tingkat dua.

Pemelajar tingkat satu cenderung menggunakan banyak nomina dan verba dasar dalam tulisan mereka, misalnya ‘jam’, ‘nama’, ‘ibu’, ‘ada’, ‘tinggal’, ‘makan’, dan ‘pergi’. Pemelajar tingkat dua hingga empat lebih bervariasi dalam penggunaan nomina dan verba. Hal ini berhubungan dengan topik atau tema tulisan pemelajar. Tulisan pemelajar tingkat satu kebanyakan tentang perkenalan diri dan keluarga, sedangkan tulisan pemelajar tingkat dua ke atas topiknya lebih bervariasi. Kecenderungan pemilihan verba dan nomina ini juga dapat dilihat dari Gambar 1. Persentase kelas kata nomina (46,2%) dan verba (17,5%) pemelajar tingkat satu lebih besar daripada persentase kelas kata nomina dan verba pemelajar tingkat dua ke atas. Pemelajar tingkat dua ke atas lebih banyak menggunakan preposisi dan konjungsi, serta adjektiva

dan adverbia dibandingkan pemelajar tingkat satu. Mengenai penggunaan pronomina dan determina, tidak ada perubahan yang signifikan antara pemelajar tingkat satu, dua, tiga, dan empat.

Pemilihan kata dan kelas kata dalam tulisan pemelajar ini tidak lepas dari topik tulisan pemelajar. Pemelajar tingkat satu yang topik tulisannya kebanyakan tentang perkenalan diri, keluarga, teman, rumah cenderung menggunakan banyak nomina, seperti nama orang, nama tempat, kata 'nama', dan kata-kata yang berhubungan dengan anggota keluarga, misalnya 'ibu', 'ayah', 'kakak', dan 'adik'. Pemelajar tingkat dua ke atas yang tulisannya mengenai kesan dan pendapat pribadi, naskah pidato, dan pendahuluan makalah cenderung menggunakan lebih banyak preposisi dan konjungsi serta adjektiva dan adverbia. Selain itu, pemilihan nomina dan verbanya pun lebih bervariasi.



Gambar 1. Persentase Kelas Kata 100 Kata Berfrekuensi Tinggi Pemelajar Tingkat Satu (kiri atas), Tingkat Dua (kanan atas), Tingkat Tiga (kiri bawah), dan Tingkat Empat (kanan bawah) (v=verba, n=nomina, adj=adjektiva, adv=adverbia, pron=pronomina, prep=preposisi, det=determina, conj=konjungsi)

4. Simpulan

Penulis menggunakan Korpus Pemelajar Bahasa Indonesia Beranotasi Eror (Koper) untuk menganalisis kecenderungan penggunaan kosakata dalam tulisan pemelajar BIPA di tujuh universitas di Jepang selama April 2023 hingga Maret 2024. Kata berfrekuensi tinggi dalam tulisan pemelajar BIPA memiliki kesamaan dengan kata berfrekuensi tinggi dalam korpus bahasa Indonesia secara umum dan korpus percakapan BIPA. Tulisan pemelajar BIPA memiliki ciri khas ragam tulis formal dan banyak menggunakan pronomina persona pertama tunggal 'saya'. Hal ini tidak lepas dari tema atau topik tulisan pemelajar BIPA yang kebanyakan mengenai perkenalan diri, keluarga, kesan dan opini pribadi, dan sebagainya. Pemelajar BIPA pemula (tingkat satu) lebih banyak menggunakan nomina dan verba dasar

dalam tulisannya, sedangkan pemelajar BIPA tingkat menengah hingga lanjut (tingkat dua ke atas) menggunakan lebih banyak variasi nomina dan verba, serta lebih banyak konjungsi (terutama konjungsi subordinatif ‘karena’), preposisi, pronomina relatif ‘yang’, pronomina/determina ‘itu’, dan verba kopula ‘adalah’. Selain karena meningkatnya jumlah dan variasi kosakata yang dipelajari oleh pemelajar BIPA tingkat dua ke atas, tema atau topik tulisan juga memengaruhi penggunaan kosakata dalam tulisan pemelajar. Pemelajar tingkat satu lebih banyak menulis tentang perkenalan diri, keluarga, teman, dan rumah yang cenderung lebih banyak menggunakan nomina seperti nama orang dan tempat serta kosakata yang berhubungan dengan anggota keluarga. Pemelajar tingkat dua ke atas lebih banyak menulis tentang kesan dan opini pribadi sehingga konjungsi dan preposisi serta variasi nomina, verba, adjektiva, dan adverbia lebih banyak digunakan. Hasil penelitian ini diharapkan dapat berguna bagi pengajaran BIPA dan pengembangan materi ajar BIPA.

Ucapan Terima Kasih

Terima kasih kepada semua kontributor data, anotator, penginput data, dan tim teknis yang telah membantu dalam penelitian ini. Penelitian ini didanai oleh JSPS Kakenhi Nomor 23K12235 (April 2023—Maret 2026).

Daftar Pustaka

Anthony, L. (2024). AntConc (Versi 4.3.1) [Perangkat lunak]. https://www.laurenceanthony.net/software/antconc/

Eckart, T. dan U. Quasthoff. (2013) Statistical Corpus and Language Comparison on Comparable Corpora. Dalam *Building and Using Comparable Corpora*. Springer-Verlag Berlin Heidelberg.

Markus, I.M., Hananto, dan Kurnia, N. (2011). *Vocabulary in BIPA course-books and the frequency-based Indonesian word lists*. Atma Jaya Conference on Corpus Studies (Concorps). Jakarta: Unika Atma Jaya.

Muhadjir dkk. (1996). *Frekuensi Kosakata Bahasa Indonesia*. Depok: Fakultas Sastra Universitas Indonesia.

Purwiyanti, Y., Suwandi, S., dan Andayani, N. (2017). Strategi Komunikasi Pemelajar Bahasa Indonesia bagi Penutur Asing Asal Filipina. *Ranah: Jurnal Kajian Bahasa*, 6(2), 160.

Sánchez Gutiérrez, C. H., Serrano, M. P., dan García, P. R. (2019). The effects of word frequency and typographical enhancement on incidental vocabulary learning in reading. *Journal of Spanish Language Teaching*, 6(1), 14-31.

Siagian, E. N. (2020). Kata Berfrekuensi Tinggi dalam Pembelajaran BIPA Pemula. *Ranah: Jurnal Kajian Bahasa*, 9(2), 188-201.