

Komparasi Frekuensi Kemunculan Bentuk Standar dan Bentuk Tidak Standar Bahasa Indonesia Berbasis Korpus

Rahmad Hidayat^{1*}, Mochammad Asyar¹, Hasanuddin Chaer¹, Agusman¹,
Mohammad Fakhri Rahman Setiawan Putra¹

¹Universitas Mataram, Mataram, Indonesia

Article info

Article history:
Received: 04-11-2024
Revised : 10-02-2025
Accepted: 06-04-2025

Kata kunci:
bahasa Indonesia;
bentuk standar dan
tidak standar;
linguistik korpus

Keywords:
corpus linguistics;
Indonesian;
standard and
nonstandard

A B S T R A C T

Penelitian ini bertujuan untuk mengungkapkan gambaran frekuensi kemunculan bentuk standar dan tidak standar dalam korpus. Korpus yang digunakan adalah korpus LCC Indonesian 2023 yang terdapat dalam situs web <https://cqpweb.lancs.ac.uk/>. Korpus ini mencakup berbagai teks dari tahun 2008 sampai dengan 2022. Data yang diperbandingkan dalam penelitian ini terdiri dari 12 kategori yang secara keseluruhan terdiri dari 150 bentuk. Analisis data dilakukan dengan membandingkan angka frekuensi antara bentuk standar dan tidak standar. Hasil penelitian mengungkapkan bahwa penggunaan bentuk standar dalam korpus lebih banyak dibandingkan dengan bentuk tidak standar, yakni 76,46% berbanding 23,44%. Ditemukan pula jumlah bentuk tidak standar lebih banyak dibandingkan bentuk standar pada kategori tertentu serta masih terdapat penggunaan bentuk lewah/mubazir. Penggunaan bentuk tidak standar masih dilakukan oleh penutur bahasa Indonesia sehingga diperlukan langkah strategis untuk mereduksinya.

Comparison of Occurrence Frequency of Standard and Nonstandard Forms of Corpus-Based Indonesian Language

This study aims to describe the frequency of occurrence of standard forms and nonstandard forms in the corpus. The Indonesian 2023 LCC corpus is found on the website <https://cqpweb.lancs.ac.uk/>. This corpus includes various texts from 2008 to 2022. The data compared in this study consisted of 12 categories, which in total consisted of 150 forms. The data was analyzed by comparing the frequency rate between standard and nonstandard forms. The results revealed that the use of standard forms in the corpus was more than the nonstandard forms, 76.46% versus 23.44%. It was also found that the number of nonstandard forms is more than the standard forms in certain categories, and redundant forms are still used. It is necessary to take strategic measures to decrease the use of nonstandard forms by Indonesian speakers.

Copyright © 2025 Universitas Islam Negeri Siber Syekh Nurjati Cirebon.

All rights reserved..

Corresponding author: Rahmad Hidayat, Universitas Mataram, Mataram, Indonesia

E-mail address: rahmad_ab@unram.ac.id

PENDAHULUAN

Pengutamaan bahasa Indonesia di ruang publik merupakan langkah strategis Badan Pengembangan dan Pembinaan Bahasa (Badan Bahasa) dalam menjaga bahasa Indonesia sebagai identitas keindonesiaan di wilayah publik. Salah satu indikator kesuksesan pembinaan bahasa Indonesia di ruang publik yakni ditunjukkan oleh peningkatan sikap positif penutur bahasa Indonesia terhadap bahasa Indonesia. Sikap positif yang dimaksud adalah kesadaran masyarakat akan norma bahasa Indonesia dalam praktik berbahasanya (Prayitno et al., 2022). Badan Bahasa telah berupaya memasyarakatkan bahasa Indonesia melalui berbagai strategi dan metode,

baik membangun kerja sama dengan berbagai lembaga maupun melakukan sosialisasi melalui berbagai media (Audrey, Sari & Mubarak, 2023).

Salah satu langkah evaluatif yang dapat dilakukan yakni melihat kenyataan berbahasa penutur bahasa Indonesia dalam korpus (Bargat, 2023). Langkah evaluatif ini perlu direalisasikan untuk mengetahui posisi pencapaian program berdasarkan fenomena kebahasaan secara nyata di masyarakat khususnya melalui penelitian (Norris, 2016). Penelitian ini berfokus melihat kenyataan berbahasa penutur berupa penggunaan bentuk standar dan bentuk tidak standar. Penggunaan bentuk standar yang lebih banyak dibandingkan bentuk tidak standar menjadi salah satu indikasi keberhasilan program.

Selama ini, Badan Bahasa telah melakukan berbagai kegiatan penyuluhan melalui berbagai media. Salah satu media penyuluhan yang telah dipublikasikan adalah buku Seri Penyuluhan Bahasa Indonesia yang mencakup beberapa variabel tata bahasa Indonesia (Mustakim, 2019; Sasangka, 2019; Sriyanto, 2019; Suladi, 2019). Buku ini berisi salah kaprah dan perbaikannya dalam kenyataan berbahasa penutur bahasa Indonesia. Isi buku ini sangat cocok untuk dijadikan pedoman penilaian penggunaan bentuk standar dan bentuk tidak standar yang menjadi objek penelitian ini. Secara tidak langsung, akan dapat diketahui bahwa salah kaprah yang terdapat dalam buku tersebut masih dipraktikkan dalam fenomena kebahasaan atau tidak. Dengan demikian, standar penilaian akan lebih terukur.

Frekuensi penggunaan bentuk tidak standar jauh lebih dominan daripada bentuk standarnya. Kenyataan tersebut merupakan kenyataan berbahasa Indonesia yang harus mendapat perhatian khusus di tengah upaya pembinaan dan pengembangan bahasa Indonesia, pemartabatan bahasa Indonesia, terutama internasionalisasi bahasa Indonesia. Beberapa salah kaprah yang masih banyak dilakukan oleh penutur jati bahasa Indonesia dalam realitas kebahasaan ditampilkan dalam Tabel 1.

Tabel 1 Bentuk Standar dan Tidak Standar

Bentuk Standar	Bentuk Tidak Standar
salat	shalat/sholat
ustaz	ustadz
beri tahukan	beritahukan
semipermanen	semi permanen
darma	dharma

Penelitian ini merupakan penelitian yang menggunakan linguistik korpus sebagai metode utama dalam menguak data penelitian (Astri et al., 2024). Linguistik korpus merupakan kajian yang memanfaatkan kenyataan berbahasa alami penutur bahasa baik lisan maupun tulisan yang tersimpan dalam direktori diska komputer dengan ekstensi tertentu (Rajeg, 2020; Suhardijanto & Dinakaramani, 2018). Linguistik korpus digunakan karena terdiri dari lautan teks yang empiris dan sangat merepresentasikan kenyataan berbahasa di lapangan (Jensen, 2014). Kenyataan berbahasa ini merupakan pemakaian bahasa alami manusia yang apa adanya sesuai dengan kompetensi dan performansinya masing-masing. Oleh sebab itu, pemanfaatan korpus sangat cocok untuk menguak realitas kebahasaan dalam fenomena berbahasa alami dan empiris manusia khususnya dalam konteks persoalan dalam penelitian ini (Flowerdew, 2017).

Penelitian terkait komparasi bentuk standar dan bentuk tidak standar telah dilakukan dalam berbagai bahasa di dunia termasuk bahasa-bahasa di Kawasan

Asia Tenggara. Penelitian-penelitian yang dimaksud seperti penelitian terkait bahasa di Malaysia (Collins, 2013), Brunei Darussalam (Svalberg, 2002), Thailand (Smalley, 1994). Penelitian terkait perihal yang sama juga sudah dilakukan di beberapa negara seperti di Eropa (Walsh, 2021), Arab (Watson, 2022), Jepang (Aoki et al., 2017), dan Afrika (Calteaux, 1996). Perbandingan bahasa standar dan nonstandar tersebut membahas soal variasi bahasa, standardisasi bahasa, hingga kondisi bahasa standar di tengah situasi multibahasa. Dalam pada itu, penelitian ini hadir dalam konteks Indonesia dengan berbasis korpus dalam skala besar. Secara khusus, isu yang dibahas berkaitan dengan standardisasi bahasa Indonesia beserta implikasinya seperti pemertabatan bahasa Indonesia dan internasionalisasi bahasa Indonesia.

Setakat kini, sepanjang penelusuran dalam konteks Indonesia, penelitian mengenai komparasi frekuensi penggunaan bentuk standar dan tidak standar yang berbasis linguistik korpus belum pernah dilakukan. Namun, penelitian yang mengusung studi komparasi tersendiri dan frekuensi satuan lingual tersendiri yang berbasis korpus sudah pernah ada. Pengkajian sebelumnya yang berkaitan erat dengan penelitian ini ialah penelitian yang dilakukan oleh Yuliatwati (2018). Tujuan penelitian tersebut adalah membandingkan penggunaan dua kata wanita dan perempuan dalam konteks, mulai dari tingkat frekuensi sampai dengan makna kedua kata yang diperbandingkan berdasarkan kolokasi dalam konteks. Penelitian Yuliatwati (2018) ini juga hampir sama prinsipnya dengan penelitian Salsabila, Yuliatwati, & Darmayanti, (2023) yang membandingkan frekuensi dua preposisi dalam bahasa Indonesia, yakni pada dan kepada. Pembandingan juga dilakukan untuk mengetahui peran sintaksis antara dua preposisi tersebut. Dapat dikatakan bahwa komparasi yang dilakukan hanya terbatas pada dua kata.

Penelitian selanjutnya dilakukan oleh Wahyuningtyas & Kesuma (2021) hanya berfokus pada identifikasi frekuensi kata yang dipergunakan untuk bahan ajar Bahasa Indonesia untuk Penutur Asing (BIPA) 1. Pengidentifikasian frekuensi kata bertujuan untuk mengetahui kata mana saja yang paling banyak digunakan dalam buku tersebut untuk menjadi bahan penyusunan kamus BIPA sehingga memudahkan penutur asing dalam belajar kata bahasa Indonesia. Prinsip yang sama juga diusung oleh penelitian yang dilakukan oleh Syartanti (2022), Syarofi & Nugraha (2022), dan Wiedarti et al. (2022), dan Shafri et al. (2023). Keempat penelitian tersebut juga menggunakan variabel frekuensi berbasis korpus dalam menjawab rumusan masalah penelitiannya.

Pada akhirnya, dapat ditegaskan bahwa fokus penelitian-penelitian di atas hanya berkonsentrasi pada variabel frekuensi. Hal ini tentu saja berbeda dengan penelitian ini yang membandingkan frekuensi penggunaan bentuk standar dan tidak standar dengan memanfaatkan sumber korpus tahun terbaru. Data yang diperbandingkan juga jauh lebih banyak untuk mendapatkan gambaran yang komprehensif. Dengan demikian, perbedaan yang signifikan sangat terlihat dengan mengusung frekuensi sekaligus mengomparasikan frekuensi yang dimaksud. Selain itu, pengomparasian mencakup lebih banyak data yang lebih variatif dan luas. Berdasarkan hal tersebut, penelitian-penelitian yang telah disebutkan hanya memiliki kesamaan pemanfaatan korpus dengan melihat tingkat frekuensi variabel utama yang sedang diteliti. Namun, variabel komparasi, jumlah data, serta tujuan penelitian dapat dinyatakan jauh berbeda.

Pembandingan bentuk standar dengan bentuk tidak standar bahasa Indonesia berbasis korpus dalam penelitian ini hadir karena belum terdapat penelitian

komparatif yang memanfaatkan data skala besar secara holistik. Di samping itu, keterbatasan ketersediaan dan keterbatasan akses terhadap korpus sebelumnya juga memengaruhi hal tersebut. Sementara itu, gambaran representatif hanya bisa didapatkan dengan pemanfaatan data skala besar yang mencakup rentang waktu yang relatif lama dan panjang. Penelitian ini dapat memberi gambaran yang komprehensif dengan berdasar pada data skala besar yang belum pernah dilakukan sebelumnya.

Penelitian ini memiliki tujuan untuk melihat kenyataan pemakaian bahasa Indonesia di ruang publik berdasarkan gambaran frekuensi kemunculan bentuk standar dan bentuk tidak standar dalam kenyataan berbahasa. Gambaran frekuensi itu merupakan representasi kondisi pengutamaan bahasa Indonesia di lapangan. Gambaran ini selanjutnya berguna sebagai bahan evaluasi serta pembinaan dan pengembangan yang berkaitan dengan pemakaian bahasa Indonesia yang baik dan benar. Hasil penelitian juga bermanfaat bagi pembelajaran bahasa di jenjang pendidikan tertentu. Secara khusus, hasil penelitian ini dapat berguna bagi penstandaran dan pengembangan pembelajaran BIPA yang pada akhirnya mendukung internasionalisasi bahasa Indonesia.

METODE

Penelitian ini menganut pendekatan kuantitatif sekaligus kualitatif. Pendekatan kuantitatif digunakan dalam mengidentifikasi jumlah frekuensi kemunculan bentuk yang menjadi variabel utama penelitian (Han, 2020; Vessey, 2017). Pendekatan kuantitatif dimasukkan di sini karena metode linguistik korpus bersandar pada kuantitas data yang diteliti untuk selanjutnya dijabarkan secara kualitatif deskriptif berdasarkan sudut pandang teori yang digunakan (Gries, 2012; Pollach, 2012). Selanjutnya, sumber data penelitian adalah data korpus. Korpus yang digunakan adalah korpus LCC Indonesian 2023 yang terdapat dalam situs web <https://cqpweb.lancs.ac.uk/> (Hardie, 2012).

Korpus LCC Indonesian 2023 terdiri atas 43 teks, 3.471.245 tipe kata, dan 573.557.097 token. Keseluruhan teks korpus LCC Indonesian 2023 diambil dari korpus Leipzig Corpora Collection mulai tahun 2008 sampai dengan tahun 2022. Berdasarkan *corpus manual*-nya, korpus bahasa Indonesia di CQPweb berasal dari data Leipzig Corpora Collection. Korpus ini berasal dari berbagai sumber yang terdiri atas sumber berita, wikipedia, situs web publik, serta berbagai sumber lainnya (*mixed*). Oleh karena itu, korpus ini dapat dikatakan representatif dan berimbang dalam merekam realitas penggunaan bahasa Indonesia karena terdiri atas aneka sumber baik formal maupun nonformal. Dengan demikian, validitas data dapat dipertanggungjawabkan secara kuantitatif dan kualitatif.

Data dalam penelitian ini dikumpulkan menggunakan metode linguistik korpus. Linguistik korpus merupakan studi pemanfaatan korpus berikut peranti analisisnya untuk menafsirkan fenomena atau kenyataan berbahasa (McEnery & Brookes, 2024). Pengumpulan data penelitian ini dilakukan dengan mengakses laman <https://cqpweb.lancs.ac.uk/>. Data penelitian dikumpulkan dari korpus melalui *standard query* dengan menggunakan *wildcard* “(X|X)” seperti (respon|respons) agar secara langsung menghadirkan komparasi *number of occurrences* dan komparasi persentase melalui fitur *frequency breakdown*. Frekuensi yang digunakan dalam penelitian ini adalah frekuensi absolut dalam korpus (Bentz et al., 2017; Neil Bermel, 2018). Data yang diperbandingkan dalam

penelitian ini terdiri atas 12 kategori yang secara keseluruhan terdiri atas 150 bentuk. Di samping itu, terdapat pula data bentuk lewah.

Sebagaimana disebutkan di atas, data yang diperbandingkan terdiri atas 12 kategori. Kategori yang dimaksud adalah pemakaian gabungan huruf konsonan, penulisan kata majemuk berimbuhan, penulisan bentuk terikat, penulisan kata majemuk, penulisan bentuk yang sudah dianggap padu, penulisan kata depan, penulisan imbuhan di-, penulisan bilangan tingkat dan kumpulan, penulisan partikel, penulisan ke- dengan angka, pemakaian imbuhan, dan pemakaian bentuk baku dan tidak baku. Sementara itu, bentuk lewah terdiri atas 8 data.

Terkait dengan standar data yang diperbandingkan, penelitian ini memanfaatkan buku Seri Penyuluhan Bahasa Indonesia yang dipublikasikan oleh Badan Bahasa. Data yang akan diperbandingkan merupakan data salah kaprah yang sering dipraktikkan oleh masyarakat penutur bahasa Indonesia berdasarkan hasil kajian Badan Bahasa. Penggunaan data yang dimaksud adalah untuk melihat data tersebut masih menjadi salah kaprah atau tidak dengan melihat frekuensi penggunaannya di lapangan. Dengan demikian, dapat diketahui tingkat keberhasilan program Badan Bahasa selama ini jika frekuensi kemunculan bentuk standar lebih banyak dibandingkan bentuk tidak standar.

Pada tahapan penganalisisan data, data berupa *number of occurrences* dan *frequency* hasil dari *frequency breakdown* dihimpun dalam tabel perbandingan. Setelah itu, dilakukan penghitungan rata-rata *number of occurrences* dan *frequency* setiap kategori yang diperbandingkan dan rata-rata *number of occurrences* dan *frequency* secara keseluruhan (Hardie, 2012). Setelah data kuantitas bentuk standar dan bentuk tidak standar didapatkan, dilakukan pengontrasan dan pengomparasian kuantitas untuk kemudian dibandingkan dengan teori, kebijakan, dan aturan yang berlaku terkait pengutamaan bahasa Indonesia di ruang publik.

Setelah data dianalisis, data penelitian ini selanjutnya disajikan dengan menggunakan metode formal dan metode informal (Mahsun, 2019). Metode formal adalah metode penyajian hasil analisis data menggunakan simbol dan lambang tertentu. Sementara itu, metode informal adalah cara penyajian hasil analisis data dengan menggunakan uraian kata-kata sebagaimana biasanya. Seluruh data yang telah teranalisis akan dipaparkan secara deskriptif sesuai dengan rumusan masalah dan tujuan penelitian yang telah ditetapkan.

HASIL DAN PEMBAHASAN

Pada bagian ini, akan disajikan berbagai grafik dan tabel yang menunjukkan komparasi bentuk standar dan bentuk tidak standar berdasarkan beberapa sudut pandang disertai uraian penjelasnya. Pada grafik, terdapat kode kategori kelompok yang diperbandingkan. Kategori yang dimaksud diperincikan sebagai berikut. K-1 merupakan kategori pemakaian gabungan huruf konsonan. K-2 merupakan kategori penulisan kata majemuk berimbuhan. K-3 merupakan penulisan bentuk terikat. K-4 merupakan kategori penulisan kata majemuk. K-5 merupakan kategori penulisan bentuk yang sudah dianggap padu. K-6 merupakan kategori penulisan kata depan. K-7 merupakan kategori penulisan imbuhan di-. K-8 merupakan kategori penulisan bilangan tingkat dan kumpulan. K-9 merupakan kategori penulisan partikel. K-10 merupakan kategori penulisan ke- dengan angka. K-11 merupakan kategori pemakaian imbuhan. K-12 merupakan kategori pemakaian bentuk baku dan tidak baku.

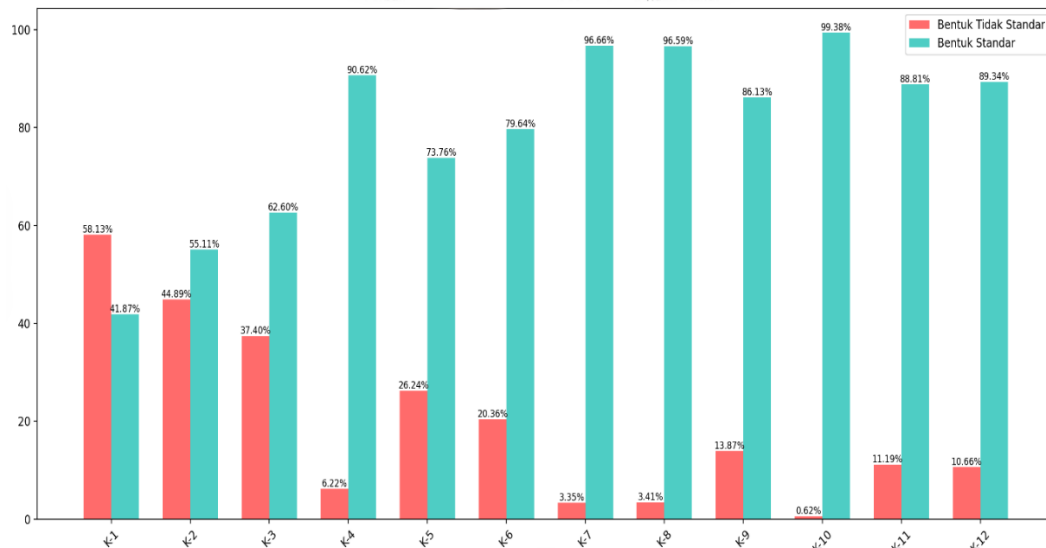
K-1 terdiri atas bentuk-bentuk seperti *hadits, lafazh, bhakti, budhi, dharma, wudhu, Ramadhan, maghrib, bathin, shalat/sholat, ashar, shubuh, ustadz, ustadzah, adzan, dan dhuha*. K-2 terdiri atas bentuk-bentuk berupa *bertandatangan, penanggungjawab, tandatangani, bekerjasama, beritahukan, bertanggungjawab, dan penandatangan*. K-3 terdiri atas bentuk-bentuk berupa *antar kota, tuna wisma, sub bagian, sub tema, non formal, manca negara, nara sumber, nara pidana, pasca banjir, pasca sarjana, semi permanen, semi formal, multi fungsi, pramu saji, dwi warna, dan dwi bahasa*. K-4 terdiri atas bentuk-bentuk berupa *tandatangan, tandamata, rumahtangga, rumahsakit, orangtua, mataacara, mataair, mejatulis, mejamakan, kakitangan, dan kakilima*. K-5 terdiri atas bentuk-bentuk berupa *acap kali, barang kali, bea siswa, darma bakti, duka cita, hulu balang, mata hari, sapu tangan, suka cita, ada kalanya, bila mana, bela sungkawa, kaca mata, mana suka, olah raga, peri bahasa, peri laku, segi tiga, wira swasta, kilo meter, kasat mata, dan suka rela*.

Selanjutnya, K-6 terdiri atas bentuk-bentuk berupa *disamping, disini, disana, disitu, diatas, dipinggir, dibawah, diantaranya, kesamping, kepinggir, kedepan, keatas, kebawah, kesini, kesana, dan kesitu*. K-7 terdiri atas bentuk-bentuk berupa *di makan, di simpan, di antar, di pinjam, di coba, dan di bawa*. K-8 terdiri atas bentuk-bentuk berupa *ke dua, ke tiga, ke empat, dan ke lima*. K-9 terdiri atas bentuk-bentuk berupa *meski pun, walau pun, sungguhpun, biar pun, kendati pun, bagaimana pun, permeter, satupersatu, dan pertelepon*. K-10 hanya terdiri atas bentuk berupa *ke dan ke-*. K-11 terdiri atas bentuk-bentuk berupa *merubah, penglepasan, pengrusak, pengrusakan, pengrajin, menyolok, memroduksi, menyontoh, menyubit, pemproses, dan pemproduksi*. K-12 terdiri atas bentuk-bentuk berupa *koordinir, publisir, legalisir, proklamir, produser, manipulir, akte, aktifitas, analisa, beaya, diagnosa, efektif, efisien, Pebruari, foto copy, hipotesa, jadwal, kualitas, kuantitas, questioner, kwitansi, metoda, Nopember, obyek, prosen, prosentase, produktifitas, Rabo, resiko, sistim, dan tehnik*.

Hasil penelitian menunjukkan bahwa secara rata-rata, penggunaan bentuk standar dalam korpus lebih banyak dibandingkan dengan bentuk tidak standar, yakni 76,46% berbanding 23,44%. Selain itu, ditemukan pula jumlah bentuk tidak standar lebih banyak dibandingkan bentuk standar pada kategori tertentu. Jumlah bentuk yang dimaksud teridentifikasi sebanyak 25 bentuk. Terdapat pula bentuk lewah yang masih teridentifikasi penggunaannya dalam korpus. Perlu disampaikan bahwa hasil penelitian ini relatif representatif karena isi korpus yang digunakan cukup seimbang mewakili penggunaan bahasa Indonesia di ragam formal maupun informal. Berdasarkan metadatanya, korpus LCC Indonesian 2023 merupakan korpus yang sumbernya dihimpun dari tahun 2008—2022. Data korpus bersumber dari berbagai ragam sumber seperti situs web, Wikipedia, berita, dan campuran dari berbagai sumber. Kerepresentatifan ini penting disampaikan karena akan memengaruhi hasil penelitian. Jika isi korpus didominasi oleh sumber yang relatif ketat dalam penulisan, penggunaan bentuk standar otomatis akan mendominasi hasil penelitian. Sebaliknya, jika isi korpus berasal dari berbagai sumber, baik sumber formal maupun informal, hasil penelitian dapat dinyatakan relatif mewakili penggunaan bentuk standar maupun bentuk tidak standar.

Komparasi Jumlah Kemunculan Bentuk Standar dan Bentuk Tidak Standar dalam LCC Indonesian 2023

Berdasarkan perhitungan rata-rata per kategori, Gambar 1 menyajikan komparasi frekuensi kemunculan bentuk standar dan bentuk tidak standar dengan persentase.



Gambar 1 Komparasi Frekuensi Kemunculan Bentuk Standar dan Bentuk Tidak Standar dalam LCC Indonesian 2023

Berdasarkan grafik, frekuensi kemunculan bentuk standar dan tidak standar dalam korpus LCC Indonesian 2023 menunjukkan bahwa sebagian besar kategori bentuk standar lebih tinggi daripada bentuk tidak standar. Hal ini terlihat dari data K-2 sampai dengan K-12 yang menunjukkan perbedaan yang relatif signifikan antara frekuensi kemunculan bentuk standar dan bentuk tidak standar. Berdasarkan grafik, hanya K-1 yang bentuk tidak standar lebih tinggi daripada bentuk standar. Sebagaimana disebutkan di atas, K-1 merupakan kategori pemakaian gabungan huruf konsonan. Tabel 2 memperlihatkan detail perbandingan rata-rata frekuensi kemunculan sekaligus angka persentase bentuk standar dan tidak standar tahun 2023 dalam masing-masing kelompok.

Berdasarkan Tabel 2, meskipun frekuensi kemunculan bentuk standar lebih tinggi daripada bentuk tidak standar, jumlah kemunculan bentuk tidak standar patut mendapat perhatian. Hal ini khususnya pada kategori atau kelompok yang bentuk tidak standar lebih tinggi daripada bentuk standar. Meskipun kecil secara perbandingan, angka ribuan kemunculan bentuk tidak standar dalam korpus menunjukkan bahwa masih relatif banyak penutur bahasa Indonesia yang masih mempraktikkan salah kaprah penggunaan bahasa Indonesia. Di sisi yang lain, terdapat pula kemunculan bentuk standar pada kategori tertentu yang telah mencapai angka di atas 90%. Berdasarkan tabel di atas, kategori yang dimaksud adalah K-4 (penulisan kata majemuk), K-7 (penulisan imbuhan di-), K-8 (penulisan bilangan tingkat dan kumpulan), dan K-9 (penulisan partikel).

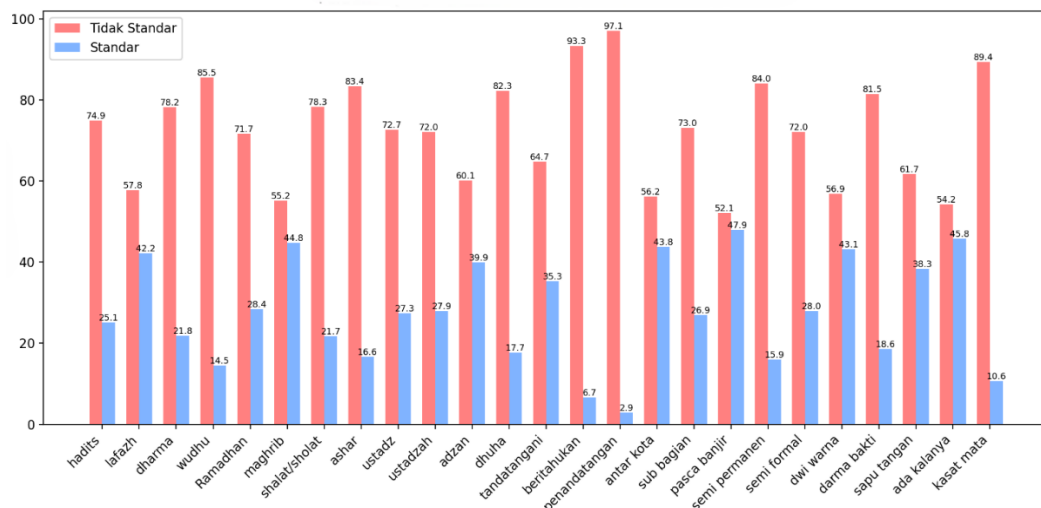
Tabel 2. Komparasi Frekuensi Kemunculan dan Persentase Bentuk Standar dan Bentuk Tidak Standar dalam LCC Indonesian 2023

Kelompok (2023)	Jumlah Kemunculan BTS	Jumlah Kemunculan BS	Persentase BTS	Persentase BS
Pemakaian gabungan huruf konsonan	9.482	8.249	58,13	41,87
Penulisan kata majemuk berimbuhan	5.398	10.851	44,89	55,11
Penulisan bentuk terikat	741	2.736	37,40	62,60
Penulisan kata majemuk	3.000	21.937	6,22	90,62
Penulisan bentuk yang sudah dianggap padu	1.107	13.405	26,24	73,76
Penulisan kata depan	13.653	42.880	20,36	79,64
Penulisan imbuhan di-	558	17.376	3,35	96,66
Penulisan bilangan tingkat dan kumpulan	3.469	180.456	3,41	96,59
Penulisan partikel	1.173	26.928	13,87	86,13
Penulisan ke- dengan angka	235	37.648	0,62	99,38
Pemakaian imbuhan	1.710	10.237	11,19	88,81
Pemakaian bentuk baku dan tidak baku	3.073	68.650	10,66	89,34

Bentuk Tidak Standar Lebih Banyak daripada Bentuk Standar

Selain teridentifikasinya kategori K-1 yang bentuk tidak standar lebih tinggi daripada bentuk standar, berdasarkan grafik di bawah ini, teridentifikasi pula beberapa bentuk dalam bank data yang menunjukkan frekuensi kemunculan bentuk tidak standar lebih tinggi daripada bentuk standar. Bentuk-bentuk yang dimaksud ada yang menjadi bagian kategori K-1 dan ada juga yang menjadi bagian dari kategori yang lain. Yang patut menjadi catatan adalah sebagian besar bentuk menunjukkan perbedaan persentase yang relatif jauh antara bentuk tidak standar dan bentuk standar. Dengan demikian, rentang perbedaan tersebut perlu mendapat perhatian untuk terus dapat diminimalisasi.

Pada Gambar 2, terdapat 25 bentuk tidak standar yang frekuensi kemunculannya lebih tinggi daripada bentuk standar. Sebagaimana dinyatakan sebelumnya, bentuk-bentuk dalam K-1 adalah bentuk yang dominan muncul. Berdasarkan grafik, selain K-1 (*hadits, lafazh, dharma, wudhu, Ramadhan, maghrib, shalat/sholat, ashar, ustadz, ustadzah, adzan, dan dhuha*), anggota kelompok yang muncul adalah K-2 (penulisan kata majemuk berimbuhan), K-3 (penulisan bentuk terikat) dan K-5 (penulisan bentuk yang sudah dianggap padu). Bentuk K-2 yang muncul adalah *tandatangani, beritahukan, dan penandatanganan*. Bentuk K-3 yang muncul adalah *antar kota, dwi warna, pasca banjir, semi formal, semi permanen, dan sub bagian*. Bentuk K-5 berupa *ada kalanya, darma bakti, kasat mata, dan sapu tangan*.



Gambar 2. Komparasi Bentuk Tidak Standar Lebih Banyak daripada Bentuk Standar dalam LCC Indonesian 2023

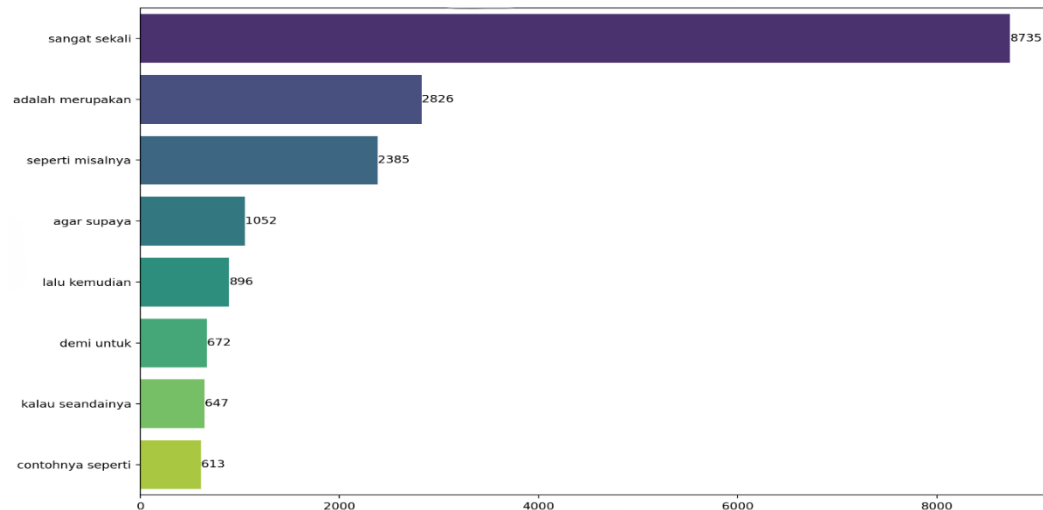
Berdasarkan Gambar 2, perlu juga disampaikan bahwa terdapat bentuk tidak standar yang angka persentase kemunculannya dalam korpus berada pada kategori 75% ke atas, bahkan berada di atas angka persentase 90%. Di samping itu, persentase bentuk tidak standar ini berjarak relatif signifikan dengan bentuk yang standar. Bentuk-bentuk yang dimaksud adalah bentuk *ashar*, *beritahukan*, *darma bakti*, *dharma*, *dhuha*, *foto copy*, *hadits*, *penandatanganan*, *Ramadhan*, *semi permanen*, *shalat/sholat*, *tandatangan*, *ustadz*, *ustadzah*, dan *wudhu*. Selain itu, dapat dilihat bahwa kelompok yang dominan dalam klasifikasi ini adalah kelompok K-1 yang merupakan kata serapan dari bahasa Arab.

Dominansi bentuk tidak standar pada kelompok serapan dari bahasa Arab dapat disebabkan oleh beberapa hal. Pertama, penutur bahasa telah terbiasa dengan pelafalan kata dalam bahasa Arab sehingga hal tersebut ditranskripsi mengikuti bunyi sebagaimana pelafalannya (Gani, 2022). Kenyataan ini juga didukung oleh adanya penghormatan terhadap konteks keagamaan sehingga terdapat semacam kekhawatiran bahwa perubahan pelafalan dapat mengubah makna dalam bahasa Arab sehingga mengabaikan bentuk standar/baku. Kedua, kebiasaan penutur dalam membaca dan melihat literatur keagamaan serta kebanyakan penulisan salah kaprah yang terjadi di kalangan penutur turut memperparah resistansi penutur menggunakan bentuk yang tidak standar/tidak baku. Salah kaprah yang terjadi dapat dijadikan standar kebenaran oleh penutur.

Berdasarkan uraian di atas, pada dasarnya, pangkal persoalan adalah kesadaran penutur jati bahasa Indonesia akan kaidah bahasa Indonesia. Oleh karena itu, temuan penelitian ini tentu harus menjadi perhatian khusus untuk ditindaklanjuti agar kondisi dominansi bentuk tidak standar dalam kategori ini dapat direduksi bahkan dihilangkan dari kenyataan berbahasa penutur jati bahasa Indonesia. Dengan demikian, Badan Pengembangan dan Pembinaan Bahasa selaku lembaga yang bertugas menjaga pengutamaan bahasa Indonesia termasuk penggunaan bahasa Indonesia yang baik dan benar di ruang publik perlu melanjutkan langkah strategis yang telah dirumuskan dan menindaklanjuti beberapa temuan penelitian ini.

Jumlah Kemunculan Bentuk Pasangan Lewah (Mubazir)

Selain bentuk-bentuk yang disampaikan di atas, di bawah ini disampaikan pula grafik frekuensi kemunculan bentuk lewah atau bentuk mubazir dalam bahasa Indonesia yang terdapat dalam korpus. Bentuk-bentuk dalam kategori ini hanya diidentifikasi jumlah kemunculannya tanpa perbandingan dengan bentuk standarnya, seperti pada Gambar 3.



Gambar 3. Kemunculan Bentuk Lewah (Mubazir) dalam LCC Indonesian 2023

Berdasarkan Gambar 3, secara keseluruhan, penggunaan bentuk lewah tertinggi adalah pasangan *sangat ... sekali* yang mencapai 8.735 kemunculan. Selain itu, bentuk-bentuk yang mencapai kemunculan di atas seribu adalah bentuk pasangan *adalah merupakan* (2.826 kemunculan), *seperti ... misalnya* (2.385 kemunculan), dan *agar ... supaya* (1.052 kemunculan). Meskipun demikian, jika dibandingkan dengan jumlah token korpus yang mencapai ratusan juta, angka tersebut masih tergolong sangat kecil. Bentuk-bentuk tersebut biasa disebut sebagai redundansi dalam bahasa. Terjadinya redundansi dalam berbahasa sangat dipengaruhi oleh fungsi praktis dalam berkomunikasi (Leufkens, 2020). Fungsi praktis yang dimaksud seperti penekanan atau penegasan makna yang ingin disampaikan serta penambahan efek emosional. Selain itu, fleksibilitas dalam berbahasa lisan juga memengaruhi kemunculan redundansi dalam berkomunikasi.

Meskipun mendukung komunikasi, kenyataan redundansi dalam bahasa juga patut menjadi catatan untuk pengambilan langkah strategis dalam rangka meminimalisasi dan mereduksi penggunaan bentuk-bentuk yang dimaksud. Realitas kebahasaan ini bertahan juga disebabkan oleh intensitas salah kaprah yang dilakukan oleh penutur bahasa khususnya dalam praktik berbahasa lisan. Dengan demikian, pangkal persoalan kembali kepada kesadaran penutur bahasa Indonesia akan kaidah kebahasaan yang berlaku dalam bahasa Indonesia. Jadi, para pemangku kepentingan diharapkan menentukan langkah tindak lanjut untuk mengatasi permasalahan yang dimaksud.

Berdasarkan hasil penelitian, frekuensi kemunculan bentuk standar lebih tinggi daripada bentuk tidak standar. Kenyataan ini dapat dibahas dari dua sudut pandang, yakni sudut pandang kebijakan Badan Bahasa dan sudut pandang linguistik yang berkaitan dengan kenyataan berbahasa. Pembahasan dari sudut

pandang kebijakan dilakukan untuk melihat keselarasan antara hasil penelitian dengan program yang telah dilaksanakan oleh Badan Bahasa. Uraian keselarasan yang dimaksud akan menunjukkan berhasilnya upaya Badan Bahasa dalam meningkatkan kualitas penggunaan bahasa Indonesia di ruang publik dan komunikasi penutur. Selanjutnya, pembahasan dari sudut pandang linguistik disampaikan untuk menjelaskan kenyataan berbahasa yang terjadi berdasarkan teori linguistik.

Berdasarkan sudut pandang kebijakan, Badan Bahasa relatif telah berhasil meningkatkan sikap positif penutur terhadap bahasa Indonesia melalui berbagai langkah strategis yang telah dirumuskan dalam rencana strategisnya. Dalam rencana strategis tahun 2015—2019, salah satu misi Badan Bahasa adalah peningkatan mutu kebahasaan dan pemakaiannya (Badan Bahasa, 2015). Dalam mendukung misi itu, Badan Bahasa melakukan peningkatan kualitas penggunaan bahasa Indonesia dalam berbagai sektor, khususnya di sektor pendidikan. Hal ini dilakukan dengan memfokuskan peningkatan kemahiran berbahasa kalangan pendidik dan tenaga kependidikan. Di samping itu, Badan Bahasa juga berusaha meningkatkan mutu kebahasaan melalui dukungan terhadap peningkatan literasi untuk menaikkan skor PISA dan peningkatan standar kemahiran berbahasa.

Berdasarkan rencana strategis tahun 2020—2024, salah satu misi Badan Bahasa yang sangat berkaitan erat dengan penggunaan bahasa Indonesia di masyarakat adalah misi meningkatkan literasi kebahasaan dan kesastraan (Badan Bahasa, 2022a). Hal ini diturunkan dalam salah satu program prioritas, yakni penguatan literasi kebahasaan dan kesastraan melalui penyusunan bahan literasi dan pembinaan penutur jati bahasa Indonesia. Adapun salah satu sasaran programnya adalah kemahiran berbahasa Indonesia. Hal tersebut dicapai melalui peningkatan jumlah penutur bahasa Indonesia yang teruji kemahirannya dan penguatan penggunaan bahasa Indonesia di ruang publik.

Pencapaian Badan Bahasa relatif sejalan dengan target Rencana Pembangunan Jangka Panjang Menengah Nasional (RPJMN), baik 2015—2019 maupun 2020—2024. Hal ini disebabkan oleh rencana strategis yang disusun oleh Badan Bahasa mengikuti rencana strategis yang dirumuskan oleh Kementerian Pendidikan dan Kebudayaan sebagai induk organisasinya. Sementara itu, rencana strategis kementerian dan lembaga negara wajib diselaraskan dengan RPJMN. Oleh karena itu, target dan pencapaian target yang terdapat dalam rencana strategis dan laporan kinerja tahunan Badan Bahasa telah merepresentasikan target RPJMN, khususnya pada bidang pendidikan, bahasa, sastra, dan kebudayaan.

Kegiatan-kegiatan yang dilaksanakan Badan Bahasa sebagai implementasi peningkatan literasi kebahasaan dan kesastraan adalah pengayaan bahan ajar literasi, pembinaan bahasa dan sastra, dan pengujian bahasa. Beberapa kegiatan dan produk yang dihasilkan sebagai perwujudan implementasi itu seperti penyusunan dan pengembangan kamus dan pedoman (KBBI, tesaurus, glosarium, ensiklopedia, EYD, PUI, dan sebagainya), Gerakan Literasi Nasional (pengadaan dan penyebaran bahan bacaan literasi), Uji Kemahiran Berbahasa Indonesia (UKBI), pembinaan bahasa Indonesia di ruang publik, pembelajaran Bahasa Indonesia untuk Penutur Asing (BIPA), penghargaan dan apresiasi untuk pengembangan bahasa dan sastra, pemantauan dan evaluasi pemakaian bahasa di media massa, penyuluhan bahasa Indonesia (pelaksanaan kursus dan penyuluhan, buku Seri Penyuluhan Bahasa Indonesia, dan sebagainya), serta pengembangan teknologi kebahasaan

(Sipebi, Padanan Istilah, Data Pokok Kebahasaan, Korpus Indonesia, Layanan Ahli Bahasa, dan sebagainya).

Upaya Badan Bahasa dalam hal peningkatan kualitas penggunaan bahasa Indonesia relatif berhasil. Pada sasaran peningkatan literasi kebahasaan dan kesastraan, terdapat dua fokus indikator kinerja program, yakni persentase penutur bahasa Indonesia terbina dan lembaga terbina yang meningkat kualitas penggunaan bahasanya. Dalam laporan kinerja Badan Bahasa tahun 2022 dan tahun 2023 (Badan Bahasa, 2022b; Badan Bahasa, 2023), berkaitan dengan peningkatan kemahiran berbahasa penutur bahasa Indonesia, Badan Bahasa telah melampaui target pencapaian penutur bahasa Indonesia terbina dan meningkat kualitas berbahasanya. Pelampauan tersebut terealisasi pada angka 68,01% dari target 66% pada tahun 2022 dan pada angka 82,25% dari target 70% pada tahun 2023. Dalam pada itu, lembaga terbina juga melampaui target pencapaian, yakni terealisasi pada angka 60% dari target 59,93% pada tahun 2022 dan 75,99% dari target 75,96% pada tahun 2023.

Salah satu faktor yang memengaruhi pencapaian tersebut adalah adanya program Uji Kemahiran Berbahasa Indonesia (UKBI). Adanya UKBI Adaptif yang dapat diikuti secara daring dan kolektif mampu meningkatkan jumlah peserta selain adanya kepentingan peserta mendapatkan skor UKBI sebagai persyaratan peningkatan karier dan sebagainya. Secara tidak langsung, UKBI telah memengaruhi penutur bahasa Indonesia untuk lebih peduli dengan kaidah-kaidah bahasa Indonesia (Widia et al., 2022). Untuk mendukung kenyataan tersebut, Badan Bahasa dalam hal ini memublikasikan superaplikasi Halo Bahasa yang berisi berbagai subaplikasi yang membantu pengguna dalam mengakses seputar kaidah bahasa Indonesia serta penggunaan bahasa Indonesia yang baik dan benar. Adanya fitur-fitur layanan kebahasaan dan kesastraan dalam superaplikasi tersebut sangat membantu pengguna dalam meningkatkan kualitas berbahasa Indonesianya. Di sisi yang lain, Badan Bahasa juga memberikan apresiasi kepada lembaga pendidikan sehingga semangat dalam mendukung peningkatan kemahiran berbahasa Indonesia terus bertambah.

Faktor lain yang mendukung peningkatan kemahiran berbahasa Indonesia adalah adanya kegiatan penyuluhan bahasa Indonesia serta kemudahan akses terhadap informasi mengenai bahasa Indonesia (Sarmadan, 2020). Penyuluhan bahasa Indonesia dilaksanakan dalam bentuk kelas-kelas daring dan pelayanan dari ahli bahasa. Seluruh program juga didukung oleh kerja sama Badan Bahasa dengan berbagai instansi dan lembaga yang ikut menjaga martabat bahasa Indonesia di ruang publik sehingga mampu menjadi contoh dan acuan bagi penutur bahasa Indonesia dalam menggunakan bahasa Indonesia yang baik terutama benar. Dengan demikian, peningkatan kemahiran berbahasa dan kualitas penggunaan bahasa Indonesia penutur dapat dicapai.

Pemerintah melalui Badan Bahasa juga telah menetapkan berbagai regulasi berkaitan dengan penguatan dan pemertabatan bahasa Indonesia. Regulasi berupa peraturan yang dimaksud adalah Peraturan Presiden Republik Indonesia 63 Tahun 2019 tentang Penggunaan Bahasa Indonesia, Peraturan Pemerintah Republik Indonesia Nomor 57 Tahun 2014 tentang Pengembangan, Pembinaan, dan Pelindungan Bahasa dan Sastra, serta Peningkatan Fungsi Bahasa Indonesia (Sekretariat Negara, 2019), Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 42 Tahun 2018 tentang Kebijakan Nasional Kebahasaan dan Kesastraan (Kemendikbud RI, 2018), Peraturan Menteri Pendidikan dan

Kebudayaan Republik Indonesia Nomor 70 Tahun 2016 tentang Standar Kemahiran Berbahasa Indonesia (Kemendikbud RI, 2016), Peraturan Menteri Dalam Negeri Nomor 40 Tahun 2007 tentang Pedoman bagi Kepala Daerah dalam Pelestarian dan Pengembangan Bahasa Negara dan Bahasa Daerah (Kemendagri RI, 2007), serta Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia Nomor 18 Tahun 2021 tentang Pembakuan dan Kodifikasi Kaidah Bahasa Indonesia Kemendikbudristek RI, 2021). Keberadaan peraturan ini semakin memperkuat pemertabatan bahasa Indonesia di ruang komunikasi penutur bahasa Indonesia.

Berdasarkan uraian di atas, dapat dinyatakan bahwa terdapat keselarasan antara hasil penelitian dengan upaya Badan Bahasa melalui program strategis yang telah diimplementasikan. Penggunaan bentuk standar yang jauh lebih banyak daripada bentuk tidak standar dalam korpus berbanding lurus dengan laporan kinerja Badan Bahasa yang melampaui target pada tahun 2023. Hal ini dapat berarti bahwa program strategis yang dilaksanakan berpengaruh positif terhadap peningkatan kemahiran berbahasa penutur jati bahasa Indonesia. Hal ini wajar terjadi karena sasaran strategis program tidak hanya penutur bahasa Indonesia, tetapi juga lembaga dan institusi yang sangat berpengaruh terhadap kualitas penggunaan bahasa Indonesia penutur.

Selanjutnya, dari sudut pandang linguistik, fenomena penggunaan bentuk standar dan bentuk tidak standar dalam realitas kebahasaan penutur bahasa Indonesia merupakan bagian dari fenomena sikap bahasa penutur bahasa (Vari & Tamburelli, 2023). Sikap bahasa merupakan sikap yang merujuk persepsi dan pandangan penutur bahasa terhadap bahasanya (Baker, 1992). Sikap ini ditunjukkan dengan kecenderungan merespons positif atau negatif terhadap bahasa. Terdapat tiga komponen sikap yang sangat berkaitan dengan sikap bahasa, yakni komponen kognitif, komponen afektif, dan komponen tindakan. Ketiganya merupakan fondasi yang berhubungan erat dan menyatu menjadi sebuah konstruksi yang menunjukkan sikap bahasa pada tingkat yang abstrak (Baker, 1992).

Penutur bahasa menilai bahasa tertentu berdasarkan sudut pandangnya ketika digunakan dalam berkomunikasi (Li & Wei, 2022). Penilaian tersebut biasanya ditimbang dari segi prestise, gengsi, kelayakan, dan sebagainya. Pada akhirnya, penutur bahasa dapat bersikap positif, negatif, atau ambivalen (netral) terhadap bahasanya. Sikap positif ditunjukkan dengan penerimaan dan penggunaan bahasa tertentu sebagai sebuah kebanggaan dan kebutuhan. Sikap negatif ditunjukkan dengan menghindari penggunaan bahasa tertentu karena dianggap kurang berprestise dan bergengsi bagi penggunaannya (Satraki, 2019). Sementara itu, sikap ambivalen (netral) cenderung fleksibel dan tidak terlalu kuat kecondongannya terhadap ragam bahasa tertentu (Menggo & Suastra, 2020).

Penggunaan bentuk standar yang dominan dalam penelitian ini menunjukkan bahwa sikap penutur bahasa Indonesia cenderung positif terhadap bahasanya. Sikap positif ini cenderung dipengaruhi oleh keyakinan kognitif bahwa berbahasa Indonesia yang benar merupakan keharusan demi penguatan identitas nasional serta kekuatan perasaan bahwa penggunaan bahasa Indonesia yang benar merupakan kebanggaan. Pada akhirnya, keyakinan itu tecermin dalam perilaku atau tindakan berupa pembiasaan berbahasa Indonesia yang benar dalam kehidupan sehari-hari (Baker, 1992). Kenyataan sikap positif ini juga akan berimplikasi pada kebijakan bahasa yang juga positif ke arah pemertabatan dan pelestarian bahasa Indonesia di masa yang akan datang.

Kebijakan bahasa yang positif sebagai implikasi dari sikap positif bahasa Indonesia adalah internasionalisasi bahasa Indonesia. Sikap positif penutur akan berdampak pada penguatan bahasa Indonesia sebagai identitas budaya. Kuatnya identitas budaya ini selanjutnya akan mendorong dan memperluas pengaruh bahasa Indonesia di kancah internasional sehingga menarik minat orang asing untuk belajar bahasa Indonesia (Rachman, Andayani, & Suyitno, 2019; Tiawati et al., 2024). Sikap positif penutur jati terhadap bahasa Indonesia juga akan membuka ruang-ruang promosi serta membangun jejaring atau komunitas belajar bahasa Indonesia di luar negeri. Dengan demikian, sikap positif penutur jati bahasa Indonesia sangat berdampak signifikan terhadap upaya internasionalisasi bahasa Indonesia yang efektif dan berkelanjutan (Sudikan, 2022).

Sikap positif dalam bentuk ketaatan terhadap kaidah bahasa juga akan mendukung upaya internasionalisasi bahasa Indonesia. Ketaatan terhadap kaidah yang di performatif secara teratur akan meningkatkan kredibilitas bahasa Indonesia sebagai bahasa yang ajek sehingga menarik minat orang asing untuk mempelajarinya karena dianggap sebagai sebuah kemudahan dalam belajar bahasa. Secara keseluruhan, ketaatan penutur bahasa Indonesia dalam menggunakan kaidah yang baku memperkuat upaya internasionalisasi bahasa Indonesia dengan menciptakan citra bahasa yang profesional, jelas, dan mudah dipelajari (Lopez, 2019). Ketaatan ini membuat bahasa Indonesia semakin dikenal dan dihargai, membuka peluang bagi bahasa Indonesia untuk lebih diterima dalam ranah akademik, diplomasi, dan sosial budaya internasional.

Selanjutnya, keberadaan berbagai ragam yang berbeda satu sama lain dalam bahasa Indonesia menandakan kekayaan bahasa Indonesia. Hal ini juga menunjukkan bahwa bahasa Indonesia memiliki fleksibilitas dalam penggunaannya. Namun demikian, dari sudut pandang preskriptif, keberadaan ragam informal relatif memengaruhi penggunaan ragam formal dalam realitas kebahasaan. Hal ini akan terjadi apabila ragam informal digunakan bersamaan dalam situasi kebahasaan ragam formal. Penggunaan itu bisa saja disebabkan oleh beberapa faktor seperti resistansi penutur bahasa Indonesia pada salah kaprah penggunaan bahasa Indonesia, kurangnya pemahaman penutur bahasa Indonesia terhadap kaidah bahasa Indonesia, penutur belum mendapat paparan kaidah bahasa Indonesia, adanya produktivitas dan kreativitas penutur bahasa Indonesia dalam menginovasi bentuk-bentuk potensial, kompleksitas aturan bahasa Indonesia, serta praktik perubahan atau perkembangan kaidah dari waktu ke waktu (Akmaluddin, 2016; Jumanto, 2014; Wibowo et al., 2020; Wijana, 2022).

Berbagai faktor yang memengaruhi masih adanya penggunaan bentuk tidak standar secara umum disebabkan oleh adanya standar kebenaran di benak penutur bahwa semakin banyak orang/lembaga/institusi yang menggunakan bentuk tidak standar tersebut maka bentuk itulah yang dianggap sebagai bentuk yang benar. Atas dasar itu, penutur bahasa Indonesia bertahan menggunakan bentuk tersebut dan cenderung resistan terhadap bentuk standar sehingga menyebabkan salah kaprah dalam berbahasa Indonesia (Wibowo et al., 2020). Hal ini diperparah dengan kurangnya pemahaman penutur bahasa Indonesia terkait kaidah bahasa Indonesia akibat belum mendapat paparan kaidah bahasa Indonesia. Di sisi yang lain, karena adanya sifat produktif dan kreatif dalam dirinya, penutur bahasa Indonesia juga relatif menciptakan bentuk-bentuk potensial dalam bahasa Indonesia. Bentuk tersebut merupakan bentuk yang belum ada dalam realitas kebahasaan, tetapi memenuhi kaidah pembentukan kata bahasa Indonesia. Pada suatu saat ke depan,

bentuk potensial ini akan ada dalam praktik kebahasaan penutur dan ikut memengaruhi penggunaan bentuk tidak standar dalam situasi kebahasaan formal.

Selanjutnya, harus diakui pula bahwa kaidah bahasa Indonesia dapat dikatakan relatif sangat kompleks bagi penutur bahasa Indonesia. Kompleksitas kaidah tersebut berkaitan dengan aspek linguistik mulai tataran fonologi, morfologi, sintaksis, dan semantik. Pada tataran fonologi misalnya, berdasarkan hasil penelitian, kategori yang didominasi oleh bentuk tidak standar adalah kategori yang berupa kesalahan penulisan gabungan huruf konsonan. Meskipun hal tersebut diatur dalam aturan ejaan, dasar kesalahannya berasal dari aspek fonotaktik dalam fonologi bahasa Indonesia. Pada tataran morfologi, kompleksitas afiks yang memiliki banyak jenis dan variasi (alomorf) juga menyulitkan penutur bahasa Indonesia dalam menggunakan afiks yang tepat untuk mewakili konsep yang ingin disampaikan. Pada tataran sintaksis, kelemahan pengetahuan mengenai struktur fungsi kalimat sering kali menjadi penyebab kesalahan kalimat yang disusun. Selain itu, berbagai jenis konjungsi dengan fungsinya masing-masing juga cukup menyulitkan penutur dalam merangkai kalimat yang berterima. Selanjutnya, pada tataran semantik, kompleksitas semantik relasional dan perbedaan komponen makna bentuk juga cukup menyulitkan penutur bahasa Indonesia dalam hal pemilihan kata untuk menyampaikan gagasannya. Kompleksitas kaidah tersebut belum lagi soal aspek pengecualian dalam kaidah.

Masih adanya penggunaan bentuk tidak standar dalam bahasa Indonesia juga disebabkan oleh adanya perubahan dan perkembangan kaidah dalam bahasa Indonesia (Sudikan, 2019). Perubahan dan perkembangan ini ada yang dapat diikuti dan ada juga yang tidak secara langsung dapat diikuti oleh penutur bahasa Indonesia. Penutur yang dapat dengan cepat mengakses perubahan atau perkembangan kaidah biasanya dari kalangan akademisi atau kalangan yang berkecimpung di dunia pendidikan. Sementara itu, bagi kalangan yang lain, perlu waktu untuk mendapatkan informasi dan melakukan penyesuaian. Rentang waktu akses informasi dan penyesuaian ini tentu menyebabkan kebingungan penutur karena harus memproses beberapa aturan kaidah dalam satu waktu. Selain itu, penutur bahasa Indonesia masih sering menggunakan bentuk tidak standar disebabkan anggapan bahwa bentuk standar dirasa kaku dan tidak alami. Bagi penutur, penggunaan bentuk tidak standar dirasa lebih fleksibel dan sudah menjadi kebiasaan.

Berbagai faktor yang menyebabkan masih adanya penggunaan bentuk tidak standar di kalangan penutur bahasa Indonesia di atas perlu terus mendapat perhatian untuk diminimalisasi dan direduksi. Hal ini penting karena penggunaan bentuk standar pada situasi formal merupakan keharusan (Satraki, 2019). Usaha yang dapat dilakukan adalah dengan membangun dan memperkuat kesadaran penutur bahasa Indonesia terhadap kaidah kebahasaan melalui dukungan kepada program-program strategis yang telah disediakan oleh Badan Pengembangan dan Pembinaan Bahasa. Di sisi yang lain, penggunaan bentuk tidak standar pada situasi komunikasi informal bukan merupakan kesalahan. Kenyataan tersebut tentu saja sesuai dengan jargon penggunaan bahasa Indonesia yang baik dan benar, yakni penggunaan bahasa Indonesia yang sesuai dengan situasi komunikasi (baik) dan sesuai dengan kaidah (benar).

SIMPULAN

Penelitian ini membuktikan bahwa penggunaan bentuk standar (76,46%) lebih dominan daripada bentuk tidak standar (23,44%) dalam korpus, meskipun terdapat 25 bentuk tidak standar yang lebih banyak pada kategori tertentu dan bentuk lewah. Temuan ini sejalan dengan upaya Badan Pengembangan dan Pembinaan Bahasa dalam meningkatkan kemahiran berbahasa Indonesia, mencerminkan sikap positif penutur terhadap bahasa Indonesia yang dapat mendorong pengakuan internasional dan menarik minat belajar melalui program BIPA. Namun, penggunaan bentuk tidak standar tetap perlu diminimalkan dengan meningkatkan akses pemahaman kaidah bahasa dan penyuluhan. Rekomendasi meliputi: (1) intervensi sosialisasi kesalahan penggunaan unsur serapan bahasa Arab, (2) penguatan sosialisasi kesalahan umum dan pembuatan pangkalan data kesalahan berbahasa berbasis AI terintegrasi Sipebi, serta (3) kolaborasi Badan Bahasa dengan media massa dan lembaga terkait sebagai teladan berbahasa di ruang publik.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Rektor Universitas Mataram melalui LPPM Universitas Mataram yang telah mendanai penelitian ini dengan dana PNBP Universitas Mataram..

DAFTAR PUSTAKA

- Akmaluddin. (2016). Problematika Bahasa Indonesia Kejinian: Sebuah Analisis Kesalahan Berbahasa Indonesia Ragam Tulisan. *Mabasan*, 10(2), 63–84. <https://doi.org/10.62107/mab.v10i2.85>
- Aoki, T., Sasano, R., Takamura, H., & Okumura, M. (2017). Distinguishing Japanese Non-standard Usages from Standard Ones. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2323–2328. <https://doi.org/10.18653/v1/d17-1246>
- Astri, Z., Noni, N., Halim, A., & Noer, F. (2024). A Comprehensive Guide to Corpus Linguistics: A Book Review of Corpora in Applied Linguistics. *Research and Innovation in Applied Linguistics-Electronic Journal*, 2(2), 174–179. <https://doi.org/10.31963/rial.v2i2.4861>
- Audrey, J., Sari, D. S., & Mubarak, K. Z. (2023). Indonesian Language Diplomacy in the United States of America Through the Bahasa Indonesia bagi Penutur Asing (BIPA) Program. *Language, Education and Culture Research*, 3(2), 1–18. <https://doi.org/10.22158/lecr.v3n2p1>
- Badan Bahasa. (2022a). Laporan Kinerja Badan Pengembangan dan Pembinaan Bahasa Tahun 2022, 188.
- Badan Bahasa. (2022b). Rencana Strategis Badan Pengembangan dan Pembinaan Bahasa Tahun 2020--2024, 50.
- Badan Bahasa. (2023). Laporan Kinerja Badan Pengembangan dan Pembinaan Bahasa Tahun 2023, 370.
- Badan Bahasa. (2015) Rencana Strategis Badan Pengembangan dan Pembinaan Bahasa 2015-2019, 55 (2015).
- Baker, C. (1992). *Attitudes and Languages* (D. Sharp, Ed.). Multilingual Matters Ltd.
- Bargat, A. (2023). Language Program Evaluation. The IAFOR International Conference on Arts & Humanities – Hawaii 2023. *Official Conference Proceedings*, 331–338. <https://doi.org/10.22492/issn.2432-4604.2023.27>

- Bentz, C., Alikaniotis, D., Samardžić, T., & Buttery, P. (2017). Variation in Word Frequency Distributions: Definitions, Measures and Implications for a Corpus-Based Language Typology. *Journal of Quantitative Linguistics*, 24(2–3), 128–162. <https://doi.org/10.1080/09296174.2016.1265792>
- Calteaux, K. (1996). *Standard and Non-standard African Language Varieties in the Urban Areas of South Africa*. New York: HSRC Publisher.
- Collins, J. T. (2013). Malay Dialect Research in Malaysia: The Issue of Perspective. *Bijdragen Tot de Taal-, Land- En Volkenkunde. Journal of the Humanities and Social Sciences of Southeast Asia*, 145(2), 235–264. <https://doi.org/10.1163/22134379-90003253>
- Flowerdew, J. (2017). Corpus-based Approaches to Language Description for Specialized Academic Writing. *Language Teaching*, 50(1), 90–106. <https://doi.org/10.1017/S0261444814000378>
- Gani, R. A. (2022). Gejala Bahasa pada Penulisan Istilah Keagamaan dan Kata Serapan dari Bahasa Arab. *Refleksi*, 21(2), 263–284. <https://doi.org/10.15408/ref.v21i2.29507>
- Gries, S. T. (2012). *Corpus Linguistics: Quantitative Methods: The Encyclopedia of Applied Linguistics*. USA: John Wiley & Sons Inc. <https://doi.org/10.1002/9781405198431.wbeal0258>
- Han, C. (2020). Statistics in Corpus Linguistics: A Practical Guide. *Journal of Quantitative Linguistics*, 27(4), 379–383. <https://doi.org/10.1080/09296174.2019.1646069>
- Hardie, A. (2012). CQPweb — Combining Power, Flexibility and Usability in a Corpus Analysis Tool. *International Journal of Corpus Linguistics*, 17(3), 380–409. <https://doi.org/10.1075/ijcl.17.3.04har>
- Jensen, K. E. (2014). Linguistics in the Digital Humanities: (Computational) Corpus Linguistics. *MedieKultur: Journal of Media and Communication Research*, 30(57), 115–134. <https://doi.org/10.7146/mediekultur.v30i57.15968>
- Jumanto, J. (2014). Politeness and Camaraderie: How Types of Form Matter in Indonesian Context. *The Second International Conference on Education and Language (2nd ICEL) 2014 Bandar Lampung University (UBL)*, 335–350. https://www.academia.edu/6389834/Politeness_and_Camaraderie_How_Types_of_Form_Matter_in_Indonesian_Context
- Kemendikbud. (2016). Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 70 Tahun 2016 tentang Standar Kemahiran Berbahasa Indonesia. In Lembaran Negara RI Tahun 2016, Sekretariat Negara. Jakarta.
- Leufkens, S. (2020). A Functionalist Typology of Redundancy. *Revista Da Abralin*, 19(3), 79–103. <https://doi.org/10.25189/rabralin.v19i31722>
- Li, C., & Wei, L. (2022). Language Attitudes: Construct, Measurement, and Associations with Language Achievements. *Journal of Multilingual and Multicultural Development*, 1–26. <https://doi.org/10.1080/01434632.2022.2137516>
- Lopez, L. G. M. (2019). The Indonesian Language and its Potential to Become an International Language. *Proceedings of the Third International Conference of Arts, Language and Culture (ICALC 2018)*, 278–280. <https://doi.org/10.2991/icalc-18.2019.40>
- Mahsun, M. (2019). *Metode Penelitian Bahasa: Tahapan, Strategi, Metode, dan Tekniknya* (3rd ed.). Depok: Rajawali Pers.

- McEnery, T., & Brookes, G. (2024). Corpus Linguistics and the Social Sciences. *Corpus Linguistics and Linguistic Theory*, 20(3), 1–23. <https://doi.org/10.1515/cllt-2024-0036>
- Menggo, S., & Suastra, I. M. (2020). Language Use and Language Attitudes of Sumbawanese Speakers in Bali. *Register Journal*, 13(2), 333–350. <https://doi.org/10.18326/rgt.v13i2.333-350>
- Mustakim. (2019). *Seri Penyuluhan Bahasa Indonesia: Bentuk dan Pilihan Kata*. Jakarta: Badan Pengembangan Bahasa dan Perbukuan.
- Neil Bermel, L. K. J. R. (2018). Frequency Data from Corpora Partially Explain Native-Speaker Ratings and Choices in Overabundant Paradigm Cells. *Corpus Linguistics and Linguistic Theory*, 14(2), 197–231. <https://doi.org/https://doi.org/10.1515/cllt-2016-0032>
- Norris, J. M. (2016). Language Program Evaluation. *Modern Language Journal*, 100(S1), 169–189. <https://doi.org/10.1111/modl.12307>
- Peraturan Menteri Dalam Negeri. (2007). Nomor 40 Tahun 2007 Tentang Pedoman bagi Kepala Daerah dalam Pelestarian dan Pengembangan Bahasa Negara dan Bahasa Daerah. [http://digilib.unila.ac.id/4949/15/BAB II.pdf](http://digilib.unila.ac.id/4949/15/BAB%20II.pdf)
- Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia. (2018). Nomor 42 Tahun 2018 Tentang Kebijakan Nasional Kebahasaan dan Kesastraan. www.kemendikbud.go.id
- Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. (2021). Nomor 18 Tahun 2021 Tentang Pembakuan dan Kodifikasi Kaidah Bahasa Indonesia.
- Peraturan Presiden Republik Indonesia (2019). Nomor 63 Tahun 2019 Tentang Penggunaan Bahasa Indonesia.
- Pollach, I. (2012). Taming Textual Data: The Contribution of Corpus Linguistics to Computer-aided Text Analysis. *Organizational Research Methods*, 15(2), 263–287. <https://doi.org/10.1177/1094428111417451>
- Prayitno, H. J., Markhamah, M., Nasucha, Y., Huda, M., Ratih, K., Ubaidullah, Rohmadi, M., Boeriswati, E., & Thambu, N. (2022). Prophetic Educational Values in the Indonesian Language Textbook: Pillars of Positive Politeness and Character Education. *Heliyon*, 8(8), e10016. <https://doi.org/10.1016/j.heliyon.2022.e10016>
- Rachman, I. F., Andayani, & Suyitno. (2019). Cultural Issues in Indonesian Language Learning for Foreign Speakers. *International Journal of Educational Research Review*, 4(3), 454–460. <https://doi.org/10.24331/ijere.573951>
- Rajeg, G. P. W. (2020). Linguistik Korpus Kuantitatif dan Kajian Semantik Leksikal Sinonim Emosi Bahasa Indonesia. *Linguistik Indonesia*, 38(2), 123–150. <https://doi.org/10.26499/li.v38i2.155>
- Salsabila, F., Yuliawati, S., & Darmayanti, N. (2023). Konstruksi Preposisi Pada dan Kepada dalam Ragam Bahasa Internet: Kajian Sintaksis Berbasis Korpus. *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, 6(3), 859–870. <https://doi.org/10.30872/diglosia.v6i3.674>
- Sarmadan, S. (2020, January). The Reinforcement Strategy of Indonesian Language in Reaching Opportunities of Becoming International Language in Supporting National Sustainability. In *Proceeding of USN Kolaka-ADRI International Conference on Sustainable Coastal-Community Development*, 1, 202–212. <https://doi.org/10.31327/icusn-adri.v1i0.1168>

- Sasangka, S. S. T. W. (2019). *Seri Penyuluhan Bahasa Indonesia: Kalimat*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.
- Satraki, M. (2019). Language Attitudes: An Overview. *International Journal of Linguistics, Literature and Translation*, 2(6), 98–113. <https://doi.org/10.32996/ijlt.2019.2.6.13>
- Shafri, M. H., Bakar, H. R. A., Raup, F. S. A., & Yusof, M. A. M. (2023). Kekekrapan Penggunaan Istilah Berkaitan COVID-19 dalam Akhbar Arab: Kajian Korpus. *International Journal of Language Education and Applied Linguistics*, 13(1), 25–35. <https://doi.org/10.15282/ijleal.v13i1.9140>
- Smalley, W. A. (1994). *Linguistic Diversity and National Unity: Language Ecology in Thailand*. University of Chicago Press.
- Sriyanto. (2019). *Seri Penyuluhan Bahasa Indonesia: Ejaan*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.
- Sudikan, S. Y. (2022). Potentials, Opportunities, and Challenges of Indonesian as an International (Scientific Journal). KIBAR 2020: *Proceedings of the 1st Konferensi Internasional Berbahasa Indonesia Universitas Indraprasta*. <https://doi.org/10.4108/eai.28-10-2020.2315276>
- Suhardijanto, T., & Dinakaramani, A. (2018). Korpus Beranotasi: Ke Arah Pengembangan Korpus Bahasa-Bahasa di Indonesia. *KBI XI*. 4–18. https://repository.kemdikbud.go.id/10088/1/dokumen_makalah_1540364204.pdf
- Suladi, S. (2019). *Seri Penyuluhan Bahasa Indonesia: Paragraf*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.
- Svalberg, A. M. L. (2002). Language Standards and Language Variation in Brunei Darussalam. *Journal of Asian Pacific Communication*, 12(1), 117–141. <https://doi.org/10.1075/japc.12.1.08sva>
- Syarofi, A., & Nugraha, A. P. (2022). Analisis Frekuensi Penggunaan Istilah Keagamaan dalam Buku Ajar Bahasa Indonesia SMA Kelas X-XII: Kajian Berbasis Korpus. *Metalingua: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 7(2), 123–130. <https://doi.org/10.21107/metalingua.v7i2.16853>
- Syantanti, N. I. (2022, July). Modalitas dalam Pidato Kenegaraan Joko Widodo: Analisis Wacana Berbasis Korpus. In *Prosiding Seminar Nasional Linguistik dan Sastra*, 202–211. <https://e-journal.unmas.ac.id/index.php/semnalisa/article/view/4720>
- Tiawati, R. L., Kurnia, M., Nazriani, Anisa, W., & Harahap, S. H. (2024). Cultural Literacy in Indonesian Language Learning for Foreign Speakers (BIPA): Overcoming Barriers and Fostering Language Proficiency with Cross-Cultural Understanding Issues. *Journal of Pragmatics and Discourse Research*, 4(1), 22–31. <https://doi.org/10.51817/jpdr.v4i1.742>
- Vari, J., & Tamburelli, M. (2023). Standardisation: Bolstering Positive Attitudes Towards Endangered Language Varieties? Evidence from Implicit Attitudes. *Journal of Multilingual and Multicultural Development*, 44(6), 447–466. <https://doi.org/10.1080/01434632.2020.1829632>
- Vessey, R. (2017). Corpus Approaches to Language Ideology. *Applied Linguistics*, 38(3), 277–296. <https://doi.org/10.1093/applin/amv023>
- Wahyuningtyas, D., & Kesuma, T. M. J. (2021). Pemanfaatan Linguistik Korpus dalam Menentukan Kata Berfrekuensi Tinggi pada Buku Sahabatku Indonesia BIPA 1. *Jurnal Bahasa Indonesia bagi Penutur Asing (JBIPA)*, 3(2), 60–69. <https://doi.org/10.26499/jbipa.v3i2.4125>

- Walsh, O. (2021). Introduction: in the Shadow of the Standard. Standard Language Ideology and Attitudes Towards ‘Non-Standard’ Varieties and Usages. *Journal of Multilingual and Multicultural Development*, 42(9), 773–782. <https://doi.org/10.1080/01434632.2020.1813146>
- Watson, J. C. E. (2022). *The Phonology and Morphology of Arabic*. In Bogoslovsk Smotra. Oxford University Press.
- Wibowo, H. A., Prawiro, T. A., Ihsan, M., Aji, A. F., Prasajo, R. E., Mahendra, R., & Fitriany, S. (2020). Semi-Supervised Low-Resource Style Transfer of Indonesian Informal to Formal Language with Iterative Forward-Translation. *International Conference on Asian Language Processing, IALP 2020*, 310–315. <https://doi.org/10.1109/IALP51396.2020.9310459>
- Widia, I., Syihabuddin, S., Damaianti, V. S., & Mulyati, Y. (2022). The Model of Bipa Listening Evaluation and its Implications for the Design of Listening Evaluation. *International Journal of Education*, 15(1), 28–37. <https://doi.org/10.17509/ije.v15i1.46154>
- Wiedarti, P., Maslakhah, S., Nirwani, I. I., & Musfiroh, T. (2022). Analisis Jumlah Kata dan Frekuensi Jenis Kata dalam Buku Elektronik Jenjang PAUD Terbitan Badan Bahasa 2019. *Litera*, 21(3), 288–300. <https://doi.org/10.21831/ltr.v21i3.53350>
- Wijana, I. D. P. (2022). Informal Vocabularies in Indonesian. *Sirok Bastra*, 10(2), 163–174. <https://doi.org/10.37671/sb.v10i2.388>
- Yuliawati, S. (2018). Perempuan atau Wanita? Perbandingan Berbasis Korpus tentang Leksikon Berbias Gender. *Paradigma, Jurnal Kajian Budaya*, 8(1), 53–70. <https://doi.org/10.17510/paradigma.v8i1.227>