



**STUDI BERBASIS KORPUS DIGITAL TERHADAP KESALAHAN EJAAN
ARTIKEL TERINDEKS SINTA**

***A DIGITAL CORPUS-BASED STUDY OF SPELLING ERRORS IN SINTA
INDEXED ARTICLES***

Evi Iryani, Pratomo Widodo, Pangesti Wiedarti, Titik Sudartinah, Ahmad Taufiq
Universitas Negeri Yogyakarta
Jalan Colombo, Nomor 1, Kampus Karangmalang, Yogyakarta, Indonesia
Ponsel: 08988112891; Posel: eviiryani.2023@student.uny.ac.id

Naskah diterima tanggal: 14 Maret 2025; Direvisi akhir tanggal: 13 Oktober 2025; Disetujui tanggal: 15 Oktober 2025
DOI: <https://doi.org/10.62107/mab.v19i2.1038>

Abstrak

Produktivitas riset di Indonesia dalam beberapa tahun terakhir mengalami lonjakan yang signifikan, namun kualitas publikasi sering kali tidak sebanding dengan kuantitasnya. Kesalahan ejaan dan penggunaan bahasa yang tidak sesuai kaidah menjadi salah satu faktor yang memengaruhi efektivitas komunikasi ilmiah. Penelitian ini bertujuan mengidentifikasi kesalahan ejaan dalam artikel-artikel jurnal terindeks SINTA 5 dan 6 menggunakan pendekatan korpus digital. Dengan memanfaatkan perangkat lunak AntConc, penelitian ini menganalisis 42 artikel dari 21 jurnal yang diperoleh dari laman resmi SINTA, difokuskan pada bidang “pendidikan bahasa”. Penelitian ini menggunakan desain *explanatory sequential* dengan pendekatan campuran, yang diawali dengan analisis kuantitatif melalui fitur *Word List*, *Cluster*, *KWIC*, dan *File View* untuk mengidentifikasi frekuensi dan pola ejaan, lalu dilanjutkan dengan analisis kualitatif untuk memahami konteks dan penyebabnya. Hasil penelitian menunjukkan bahwa kategori kesalahan ejaan paling dominan ditemukan pada penulisan kata depan “di-” dan “ke-” dengan total 305 kesalahan. Selain itu, ditemukan kesalahan pada bentuk kata turunan sebanyak 14 dan unsur serapan sebanyak 2 kesalahan. Temuan ini menunjukkan perlunya perhatian lebih dari penulis terhadap kaidah bahasa untuk meningkatkan kualitas publikasi ilmiah.

Kata kunci: korpus; kesalahan ejaan; AntConc; artikel ilmiah

Abstract

Research productivity in Indonesia has seen a significant increase in recent years; however, the quality of publications often does not match their quantity. Spelling errors and the use of language that does not adhere to established norms are among the factors that affect the effectiveness of scientific communication. This study aims to identify spelling errors in journal articles indexed in SINTA 5 and 6 using a digital corpus-based approach. Utilizing AntConc software, this study analyzes 42 articles from 21 journals

obtained from the official SINTA website, focusing on the field of “language education.” The study employs an explanatory sequential design with a mixed-method approach, beginning with quantitative analysis using features such as Word List, Cluster, KWIC, and File View to identify the frequency and patterns of spelling errors, followed by qualitative analysis to understand the context and causes. The results show that the most frequent spelling errors occurred in the use of prepositions “di-” and “ke-,” with a total of 305 errors. In addition, there were 14 errors in derivative word forms and 2 errors in loanwords. These findings highlight the need for greater attention from authors to linguistic conventions in order to improve the quality of scientific publications.

Keywords: *corpus; spelling errors; AntConc; scientific articles*

1. Pendahuluan

Dalam beberapa tahun terakhir, dunia akademik Indonesia mengalami lonjakan produktivitas riset yang signifikan. Sebuah penelitian oleh Fry dkk (2023) yang bertajuk “Ranking Researchers: Evidence from Indonesia” mengungkapkan bahwa penerapan sistem peringkat berbasis skor SINTA (Science and Technology Index) Kemendikbud Ristek telah menjadi katalis utama dalam mendorong peningkatan kuantitas publikasi ilmiah di Indonesia. Hasilnya, Indonesia kini menjadi negara penghasil artikel jurnal ilmiah terbanyak kedua di Asia Tenggara (Fry dkk., 2023). Namun, lonjakan kuantitas ini sering kali tidak diiringi oleh peningkatan kualitas, sebagaimana yang ditekankan oleh Dirjen Pendidikan Tinggi, Riset, dan Teknologi, Kemendikbud Ristek, Prof. Ir. Nizam, M.Sc., DIC, Ph.D, dalam Indonesia Research Summit 2024, yang menyatakan bahwa, “Publikasi kita meskipun banyak, tapi rata-rata jurnalnya tidak jelas, jurnal predator, dan sebagainya,” (Mashabi & Kasih, 2024).

Ketimpangan antara kuantitas dan kualitas ini salah satunya dipengaruhi karena aspek kebahasaan yang digunakan. Kualitas bahasa, termasuk ejaan, merupakan salah satu indikator penting dalam penilaian kelayakan sebuah publikasi ilmiah (Awagu, 2021; Politzer-Ahles dkk., 2020). Bahasa yang digunakan dalam jurnal artikel ilmiah harus sesuai dengan kaidah kebahasaan yang sudah ditentukan dan berpedoman pada Pedoman Umum Ejaan Bahasa Indonesia (PUEBI) (Marselina, 2022), Kamus Besar Bahasa Indonesia (KBBI), maupun Tata Bahasa Baku Bahasa Indonesia (TBBB) (Rahmanian & Kuperman, 2019). Oleh karenanya, kalimat yang dianggap baik harus memiliki pilihan kata yang tepat dan tersusun sesuai dengan aturan ejaan yang telah ditentukan (Saddler, 2023). Dalam hal ini, ejaan menjadi penentu baku atau tidaknya

sebuah kalimat (Rahmanian & Kuperman, 2019). Kalimat yang tersusun dengan baik mencerminkan kesesuaian dengan kaidah ejaan, termasuk dalam pemakaian huruf, tanda baca, serta keefektifan kalimat.

Selain itu, kesalahan pemilihan kata atau diksi, penggunaan kata baku dan tidak baku, salah tik, serta penggunaan preposisi yang tidak tepat kerap menjadi isu kebahasaan. Kesalahan leksikal atau diksi tentunya berdampak pada proses komunikasi seperti timbulnya kesalahpahaman (Wachidah, 2019). Akibatnya, kesalahan ejaan tidak hanya menurunkan kredibilitas penulis dan institusi yang bersangkutan, tetapi juga dapat menghambat komunikasi ilmiah yang efektif (Nurizka, dkk., 2021).

Meskipun demikian, kualitas riset tidak semata-mata ditentukan oleh aspek teknis seperti ejaan dan tata bahasa, tetapi juga oleh faktor fundamental seperti kejelasan topik penelitian, ketepatan rumusan masalah, ketelitian metode, kedalaman analisis, dan relevansi hasil penelitian terhadap pengembangan ilmu pengetahuan maupun masyarakat. Dengan kata lain, aspek kebahasaan memang penting sebagai sarana penyampaian ilmiah, tetapi kualitas substansi penelitianlah yang menentukan nilai ilmiahnya secara keseluruhan. Oleh karena itu, peningkatan mutu bahasa harus berjalan seiring dengan penguatan metode dan orisinalitas riset agar publikasi tidak hanya layak secara linguistik, tetapi juga bermakna secara akademik.

Kesalahan ejaan dalam tulisan ilmiah sering kali tidak terdeteksi melalui pengeditan manual, sehingga penting untuk mengembangkan pendekatan berbasis data yang dapat mengidentifikasi kesalahan-kesalahan ini secara sistematis. Khan & Rasul (2020) dalam penelitiannya mengatakan perangkat lunak berbasis korpus telah menjadi bukti efektif dalam mengidentifikasi pola kesalahan linguistik ini. Selain itu, analisis berbasis korpus dapat membantu menemukan kesalahan umum dalam dokumentasi digital, termasuk jurnal ilmiah (Jusoh, dkk., 2024).

Bednarek & Carr (2021) menguraikan bahwa analisis berbasis korpus digital tidak hanya memungkinkan penggalan pola bahasa yang tersembunyi, tetapi juga menawarkan wawasan mendalam dalam penelitian komunikasi. Namun, meski teknologi korpus digital telah menjadi alat yang diakui secara global (Liang, 2023), penelitian mengenai kualitas ejaan dalam jurnal terindeks SINTA di Indonesia masih sangat terbatas. Ketiadaan korpus yang secara khusus memetakan kesalahan ejaan dalam ranah publikasi akademik Indonesia merupakan hambatan signifikan dalam

upaya peningkatan mutu kebahasaan. Abdugafurovna, dkk (2020) menekankan pentingnya membangun korpus tematik untuk mengidentifikasi pola-pola kesalahan tertentu yang dominan dalam teks akademik. Lebih jauh lagi, korpus yang terintegrasi dengan teknik visualisasi juga dapat digunakan untuk menganalisis berbagai dimensi kesalahan bahasa, termasuk kesalahan ejaan, guna menghasilkan wawasan yang lebih mendalam (Meyer, 2023; Rayson dkk., 2016). Dalam hal ini, linguistik korpus berpotensi menjadi alat strategis untuk meningkatkan standar kepenulisan ilmiah (Wahyuningtyas & Kesuma, 2021), khususnya bagi jurnal-jurnal terindeks SINTA 5 dan 6.

Pemilihan jurnal terindeks SINTA 5 dan 6 sebagai objek kajian didasarkan pada (Kemenristekdikti, 2018) tentang akreditasi jurnal ilmiah tahun 2018, pasal 6 ayat (2), akreditasi jurnal ilmiah dibagi ke dalam enam peringkat berdasarkan skor penilaian mutu. Jurnal dengan peringkat 5 dan 6 memiliki rentang nilai antara 30 dan 49, yang mengindikasikan bahwa jurnal-jurnal tersebut masih berada pada tahap pengembangan dalam berbagai aspek mutu; termasuk di dalamnya aspek kebahasaan dan ejaan. Diperkuat oleh ketentuan dalam Pasal 4 ayat (1) yang menekankan pentingnya penggunaan bahasa yang tepat dan konsisten sebagai salah satu syarat akreditasinya, yaitu:

“Menggunakan Bahasa Indonesia dan/atau bahasa resmi Perserikatan Bangsa-Bangsa; dan menjaga konsistensi gaya penulisan dan format penampilan.”
(Permenristekdikti No. 9 Tahun 2018)

Dengan demikian, jurnal-jurnal SINTA 5 dan 6 dinilai memiliki ruang penguatan dalam penerapan kaidah kebahasaan, termasuk kepatuhan terhadap Ejaan Yang Disempurnakan. Oleh karena itu, kajian berbasis korpus terhadap kesalahan ejaan pada jurnal-jurnal dalam kategori ini menjadi krusial untuk mengidentifikasi pola kesalahan yang berulang dan memberikan kontribusi dalam peningkatan mutu kebahasaan jurnal ilmiah di Indonesia.

2. Landasan Teori

Dalam penelitian berbasis linguistik dan analisis korpus digital, beberapa teori utama digunakan sebagai kerangka pemikiran dalam mengidentifikasi dan memahami kesalahan ejaan dalam publikasi ilmiah. Landasan teori ini mencakup konsep tentang

ejaan dalam bahasa Indonesia, analisis berbasis korpus, dan implikasi kesalahan ejaan dan publikasi ilmiah.

2.1 Ejaan dalam Bahasa Indonesia

Ejaan merupakan aspek fundamental dalam keterbacaan dan kejelasan sebuah teks akademik. Menurut Pedoman Umum Ejaan Bahasa Indonesia (EYD, 2023), ejaan mencakup aturan pemakaian huruf, tanda baca, serta penulisan kata dan frasa dalam bahasa Indonesia. Kesalahan ejaan dapat menyebabkan ambiguitas dalam komunikasi akademik dan mengurangi kredibilitas suatu publikasi (Nainggolan, dkk., 2024). Selain itu, Serungke, dkk., (2023), menunjukkan bahwa kualitas kebahasaan dalam tulisan akademik berkorelasi dengan pemahaman pembaca terhadap isi tulisan. Kesalahan dalam pemilihan kata, penggunaan preposisi, serta penulisan kata baku dan tidak baku dapat menghambat komunikasi ilmiah yang efektif (Serungke, dkk., 2023). Oleh karena itu, analisis terhadap kesalahan ejaan dalam jurnal akademik menjadi penting dalam meningkatkan mutu publikasi ilmiah di Indonesia.

2.2. Analisis Berbasis Korpus

Analisis berbasis korpus merupakan metode yang digunakan untuk mengidentifikasi pola-pola kebahasaan dalam teks dengan skala besar. Metode ini memungkinkan peneliti untuk menemukan kesalahan ejaan yang sering terjadi dalam tulisan akademik dengan cara yang lebih sistematis dibandingkan dengan pemeriksaan manual (Qoiriah, 2021). Penelitian menyoroti pentingnya analisis korpus dalam mendeteksi kesalahan bahasa dalam berbagai jenis teks akademik. Lebih lanjut, pengembangan korpus khusus untuk kesalahan ejaan dalam bahasa tertentu dapat membantu dalam menemukan pola-pola kesalahan dominan dan memberikan rekomendasi perbaikan (Syahputra, 2006).

2.3. Implikasi Kesalahan Ejaan dalam Publikasi Ilmiah

Kesalahan ejaan dalam publikasi ilmiah dapat berdampak pada kredibilitas akademik dan keterbacaan sebuah artikel. Penelitian menunjukkan bahwa kesalahan linguistik dalam artikel jurnal dapat mengurangi peluang publikasi di jurnal bereputasi tinggi (Nainggolan dkk., 2024). Selain itu, dalam konteks global, standar kebahasaan yang rendah dapat menghambat kolaborasi akademik internasional serta mengurangi dampak dari hasil riset yang dihasilkan.

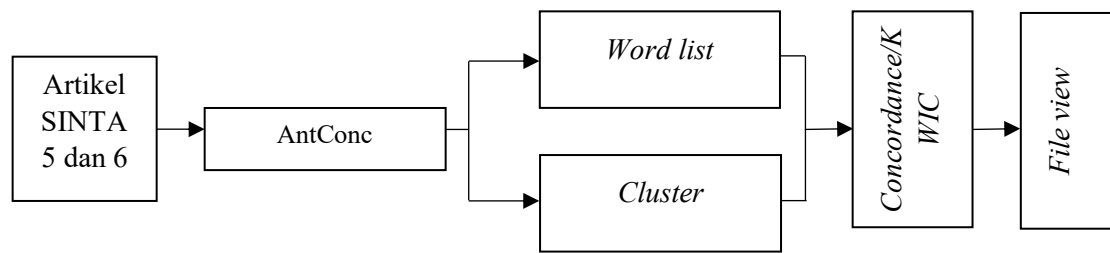
Oleh karena itu, penelitian ini akan berkontribusi dalam meningkatkan kesadaran akademisi terhadap pentingnya penggunaan bahasa yang sesuai dengan standar ejaan dalam publikasi ilmiah. Dengan menggunakan pendekatan berbasis korpus dan NLP, penelitian ini bertujuan untuk mengidentifikasi pola kesalahan ejaan dalam jurnal-jurnal SINTA 4 dan 5 serta memberikan rekomendasi konkret untuk meningkatkan kualitas publikasi ilmiah di Indonesia.

3. Metode Penelitian

Penelitian ini menggunakan pendekatan campuran (*mixed method approach*). Pendekatan ini memadukan data kualitatif dan kuantitatif sebagaimana dijelaskan oleh Creswell (2017) yang menyebutkan bahwa metode campuran memberikan pemahaman yang lebih komprehensif terhadap suatu fenomena dengan mengintegrasikan kedua jenis data.

Desain penelitian ini diawali dengan pengumpulan dan analisis data kuantitatif untuk membangun kerangka utama temuan penelitian, sebagaimana lazimnya dalam pendekatan kuantitatif yang dimulai dari perumusan pertanyaan riset dan hipotesis yang terukur (Herwanis dkk., 2025). Pendekatan ini, menurut Creswell (2017), menempatkan data kuantitatif sebagai dasar yang kokoh dalam struktur penelitian, sementara data kualitatif berfungsi untuk memberikan dimensi yang lebih mendalam terhadap temuan. Selanjutnya, data kuantitatif tersebut digunakan untuk menjelaskan, memperkaya, dan memberikan konteks terhadap hasil analisis, sebagaimana dijelaskan dalam desain *explanatory sequential* yang mengutamakan analisis data kuantitatif terlebih dahulu, yang kemudian dielaborasi melalui data kualitatif (Creswell & Plano Clark, 2017).

Sumber data dalam penelitian menggunakan artikel jurnal ilmiah yang terindeks SINTA 5 dan 6, yang diperoleh melalui lama resmi SINTA (<https://sinta.kemdikbud.go.id/journals>) yang diseleksi dengan kata kunci “pendidikan bahasa”. Artikel yang diseleksi harus ditulis dalam bahasa Indonesia, berasal dari jurnal yang terbit pada volume dan edisi terbaru (2024–2025). Pada pencarian awal ditemukan 44 jurnal dengan filter tersebut, kemudian terseleksi menjadi 21 jurnal dengan masing-masing 42 artikel. Langkah-langkah pengumpulan data dalam penelitian ini adalah sebagai berikut.



Gambar 1 Alur Pengumpulan Data

Proses pengumpulan data dimulai dari artikel-artikel yang terpilih (dalam format .pdf.) diunggah ke aplikasi AntConc. Analisis kemudian dilakukan dengan melibatkan fitur *Word List*, *cluster*, *KWIC*, dan *file view*. Pendekatan yang digunakan bersifat top-down, yang mana data disusun mengacu pada Ejaan Yang Disempurnakan (EYD edisi kelima); sehingga *Word List* digunakan untuk menampilkan daftar kata beserta frekuensinya, *Cluster* untuk menampilkan frasa yang sering muncul bersama untuk mengidentifikasi kolokasi dan pola bahasa. Dalam praktiknya, fitur cluster digunakan untuk mendeteksi penggunaan kata depan seperti “di-” dan “ke-” pada data dari EYD edisi kelima, yang kemudian divalidasi secara manual melalui *KWIC* untuk memastikan fungsinya sebagai kata depan atau imbuhan berdasarkan konteks. *KWIC* kemudian digunakan untuk menampilkan kata kunci sekaligus keabsahan konteks sekitarnya untuk menganalisis makna, penggunaan, dan pola linguistik dalam teks.

Analisis ejaan dalam penelitian ini mengikuti tahapan *spelling analysis* yang disusun berdasarkan pendekatan *rule-based corpus analysis*, dengan langkah-langkah berikut: 1) ekstraksi data awal yang dimulai dari semua kata dalam korpus diekstrak menggunakan fitur *Word List* untuk memperoleh daftar kata dan frekuensinya; 2) identifikasi kesalahan awal, dilakukan dengan menyaring berdasarkan aturan ejaan dalam EYD, terutama bentuk yang sering disalahgunakan (misalnya diatas, kedalam, dan dimana); 3) validasi konteks melalui fitur *KWIC* dengan cara memeriksa konteks kemunculannya untuk memastikan apakah kata tersebut berfungsi sebagai kata depan, imbuhan, atau bentuk turunan; 4) klasifikasi kesalahan; 5) analisis frekuensi dan konteks; 6) deskripsi kualitatif.

Tahap analisis data kualitatif dilakukan dengan meninjau langsung konteks pemakaian kata menggunakan fitur *KWIC* dan *File View*, sehingga interpretasi kesalahan tidak hanya berdasarkan bentuk, tetapi juga fungsi dan maknanya. Dataset

lengkap yang digunakan dalam analisis ini disediakan secara terpisah dan dapat diakses melalui repositori Figshare (Iryani, 2025) sebagai bentuk transparansi dan validitas penelitian.

Kemudian, untuk mendukung kejelasan temuan, hasil analisis kuantitatif akan divisualisasikan menggunakan tabel frekuensi, grafik batang distribusi kesalahan, dan cuplikan tampilan KWIC (Key Word in Context) dari perangkat AntConc. Tabel digunakan untuk menunjukkan klasifikasi kesalahan ejaan berdasarkan kategori dalam EYD, sementara visualisasi KWIC menampilkan konteks kemunculan kata secara langsung dari korpus, sehingga memungkinkan pembaca memahami bentuk kesalahan dalam konteks kalimat aslinya. Visualisasi ini bertujuan tidak hanya memperkuat validitas hasil, tetapi juga meningkatkan transparansi analisis linguistik berbasis korpus.

4. Pembahasan

Hasil temuan penelitian ini menunjukkan adanya kesalahan ejaan dalam artikel-artikel jurnal ilmiah terindeks SINTA 5 dan 6, dengan tujuan mengidentifikasi pola kesalahan berdasarkan kaidah EYD. Pembahasan pada bagian ini dibagi ke dalam beberapa subbagian yang mengulas jenis-jenis kesalahan ejaan, seberapa sering kesalahan tersebut muncul.

Tabel 1 Kategori Kesalahan Ejaan

Kategori	Jenis	Frekuensi
Penggunaan Huruf	Huruf Abjad	-
	Huruf Vokal	-
	Huruf Konsonan	-
	Gabungan Huruf Vokal	-
	Gabungan Huruf Konsonan	-
	Huruf Kapital	-
	Huruf Miring	-
	Huruf Tebal	-
Penulisan Kata	Kata Turunan	13
	Kata Depan	305
	Pemenggalan Kata	-
	Partikel	0
	Singkatan	-
	Kata Ganti	0
Penggunaan Tanda Baca	Kata Sandang	0
	Unsur Serapan	2
	Serapan Umum	2

Serapan Khusus	0
----------------	---

Tabel 1 Kategori Kesalahan Ejaan tersebut menyajikan klasifikasi kesalahan ejaan yang dianalisis melalui perangkat lunak AntConc, dengan mengacu pada kaidah EYD edisi kelima. Tanda (-) menunjukkan bahwa kategori tersebut tidak dapat dianalisis secara langsung menggunakan AntConc, sementara angka 0 menunjukkan tidak ditemukan kesalahan dalam kategori tersebut dalam korpus yang dianalisis.

Analisis korpus mengungkap kesalahan ejaan yang dikategorikan berdasarkan EYD dalam tiga kategori utama: kata turunan, kata depan, dan unsur serapan. Tujuannya untuk mengidentifikasi kesalahan yang paling sering muncul dan memahami penyebab linguistiknya. Analisis ini sejalan dengan temuan Nurizka dkk (2021) yang juga mengidentifikasi kesalahan ejaan dalam jurnal ilmiah, yang mana hasilnya memperkuat temuan ini dengan menekankan pentingnya pemahaman mendalam terhadap kaidah bahasa Indonesia guna menghindari ambiguitas makna dan efektivitas komunikasi ilmiah.

4.1 Penulisan Kata

Penulisan kata depan merupakan salah satu penting dalam kaidah ejaan Bahasa Indonesia yang baik dan benar. Kesalahan dalam penulisan kata tidak hanya berdampak pada keakuratan makna, tetapi juga memengaruhi kelayakan naskah ilmiah secara formal. Dalam EYD edisi kelima, penulisan kata mencakup berbagai bentuk, seperti kata turunan, kata depan, pemenggalan kata, partikel, singkatan, kata ganti, dan kata sandang. Analisis terhadap korpus artikel jurnal ilmiah dalam penelitian ini menunjukkan bahwa masih ditemukan berbagai kesalahan dalam penulisan kata, khususnya pada bentuk kata turunan dan gabungan kata, baik karena ketidaksesuaian dengan bentuk baku dalam Kamus Besar Bahasa Indonesia (KBBI) maupun pelanggaran terhadap kaidah penulisan gabungan kata sebagaimana tercantum dalam EYD.

Kata Turunan

Berdasarkan Ejaan Yang Disempurnakan edisi kelima yang membagi kategori kata turunan menjadi tiga jenis (kata berimbuhan, bentuk ulang, dan gabungan kata), ditemukan tiga jenis data kesalahan ejaan bahasa Indonesia. Menurut EYD, ketiga

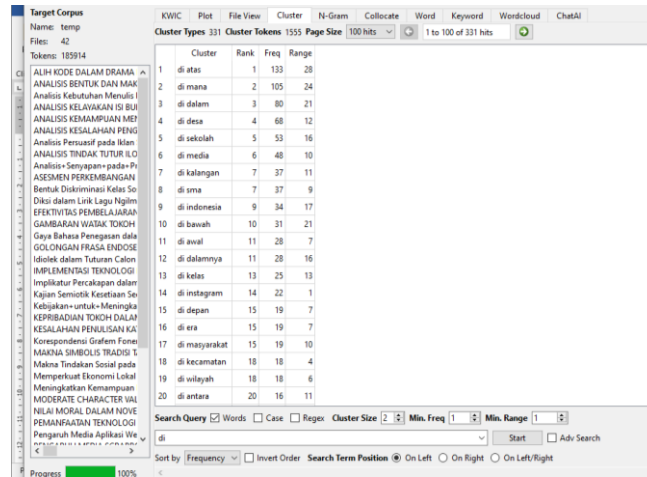
jenis tersebut secara keseluruhan tergolong dalam pasal a (unsur gabungan kata, termasuk istilah khusus, ditulis terpisah).

Dalam korpus yang dianalisis menggunakan AntConc, ketiga data tersebut, antara lain:

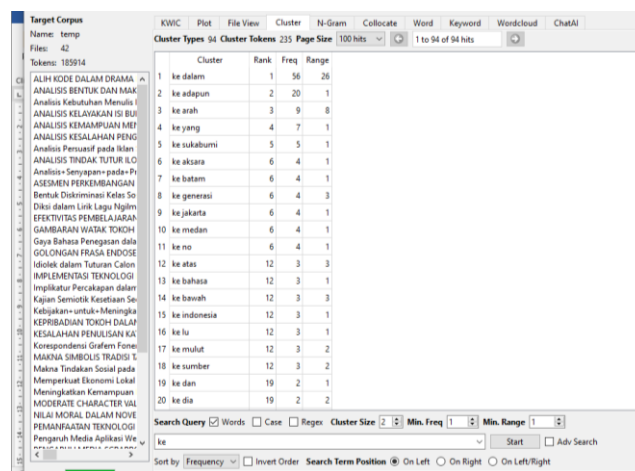
1. “Ibukota” muncul 1 kali dalam kalimat “Peristiwa mengerikan ini terjadi di Ibukota Delhi.” “Ibukota” seharusnya ditulis terpisah karena merujuk pada EYD “ibu kota” termasuk gabungan kata yang harus ditulis terpisah.
2. “Orangtua” muncul 2 kali dalam konteks seperti “dia harus memilih sendiri nerakanya, meninggalkan orangtua dan kelahirannya.” Merujuk pada EYD, penulisan “orangtua” salah karena belum mengalami pemaduan makna secara leksikal, sehingga penulisan bakunya adalah “orang tua”.
3. “kosa kata” muncul sebanyak 10 kali, seperti dalam kalimat “minimnya kosa kata bahasa Arab oleh siswa”. “Kosa kata” salah karena dalam KBBI bentuk baku yang diakui adalah “kosakata”, satu kata, sudah dianggap padu secara leksikal. Selain itu dalam EYD tentang Gabungan Kata (pasal e) dijelaskan bahwa gabungan kata berikut ditulis serangkai; dengan “kosakata” adalah salah satu contoh dari pasal tersebut.

Kata Depan

Hasil analisis korpus menggunakan fitur *cluster* (untuk mengumpulkan kemunculan “di-” dan “ke-”) dilanjutkan dengan KWIC untuk memeriksa konteks kemunculan kata, serta merujuk pedoman EYD Edisi Kelima, ditemukan sejumlah 305 kesalahan umum dalam penulisan kata depan. Namun jika merujuk pada EYD, kata depan seperti “di-” dan “ke-” harus ditulis terpisah dari kata yang mengikutinya, jika digunakan untuk menunjukkan tempat, arah, atau waktu. Oleh karena itu, fitur KWIC digunakan untuk memeriksa secara manual pada daftar kemunculan *cluster* “di-” dan “ke-” yang tidak sesuai dengan EYD.



Gambar 2 Kemunculan “di-” dalam Fitur Cluster



Gambar 3 Kemunculan “ke-” dalam Fitur Cluster

Untuk menjaga keringkas dan fokus dalam pembahasan ini, tidak semua data ditampilkan secara rinci di sini, data lengkap mengenai temuan kesalahan penulisan kata depan telah disusun dalam bentuk dataset di Figshare (tersedia dalam Iryani, 2025).

1. Penulisan kata depan “diatas” menjadi data yang cukup banyak ditemukan, yaitu sebanyak 30 frekuensi kesalahan. Kata tersebut dalam penulisannya ditulis serangkaian yang seharusnya dipisah menjadi “di atas”. Hal tersebut karena kata “di-” berfungsi sebagai kata depan dan atas merupakan kata benda. Contohnya: “Data diatas mengandung kesalahan...” yang seharusnya ditulis, “Data di atas mengandung kesalahan”.
2. Selanjutnya, kesalahan pada kata “di mana” namun ditulis dalam “dimana” ditemukan sebanyak 49 frekuensi. Kata tersebut utamanya digunakan sebagai

penghubung kalimat pernyataan, yang seharusnya digunakan sebagai bentuk pertanyaan, atau yang menunjukkan tempat serta lokasi. Misalnya ditemukan pada “...dimana tanda baca koma digunakan...”. Dalam konteks tersebut, kata di mana seharusnya ditulis menjadi di mana, atau bisa juga diganti dengan kata lain sesuai dengan konteksnya. Hal tersebut bukan hanya sekadar ejaan, tetapi juga kesalahan dalam penggunaan struktur kalimat.

3. Kata “di temukan” ditemukan sebanyak 7 frekuensi kesalahan pemisahan kata yang seharusnya digabung. Seperti pada kalimat, “... di temukan beberapa baris kalimat...”, yang mana kata tersebut merupakan kata kerja pasif dengan awalan “di-”, sehingga seharusnya ditulis serangkai; “ditemukan”, hal tersebut menunjukkan kurangnya pemahaman antara kata depan dan imbuhan utamanya “di-” dalam bahasa Indonesia.

Selanjutnya, hasil analisis korpus KWIC ditemukan sebanyak 3 kesalahan pada penulisan kata depan “ke-”.

1. Penulisan kata depan “kedalam” menjadi kesalahan yang paling sering muncul dalam kategori ini, dengan jumlah 17 frekuensi kesalahan yang tersebar di berbagai dokumen. Berdasarkan EYD edisi kelima, kata depan “ke” harus ditulis terpisah dari kata yang mengikutinya jika menunjukkan arah atau lokasi. Namun, dalam temuan korpus, bentuk “kedalam” kerap disatukan secara tidak tepat. Penulisan kata ini ditemukan dalam kalimat seperti “masuk kedalam gaya bahasa epistrofa” (dalam dokumen puisi stilistika) dan “dikelompokan kedalam dua kategori” (dokumen pembelajaran menulis). Seharusnya bentuk yang tepat adalah “ke dalam”, karena “ke” menunjukkan arah atau tempat dan “dalam” adalah nomina atau preposisi lokasi.
2. Kata “kearah” yang ditemukan sebanyak satu kali dalam korpus, yang juga merupakan kesalahan penulisan kata depan. Misalnya pada kalimat “...kearah mana saja.” Penulisan ini tidak sesuai dengan kaidah EYD, dan seharusnya ditulis “ke arah” karena “ke” berfungsi sebagai kata depan yang menunjukkan arah dan tujuan, dan “arah” adalah kata benda yang mengikutinya.
3. Kata “keindonesia” muncul 1 kali dalam korpus dan juga merupakan bentuk penulisan tidak baku. Dalam kalimat “...juga meluas keIndonesia...”, penulisan “keindonesia” menyatu secara salah. Seharusnya ditulis “ke Indonesia”, karena

menunjukkan arah atau tujuan ke suatu negara, bukan sebagai bentuk kata turunan atau imbuhan. Kesalah ini mencerminkan kurangnya pemahaman dalam membedakan kata depan dan bentuk afiksasi.

4.2 Unsur Serapan

Analisis korpus pada unsur serapan ditemukan dua kasus kesalahan dalam kategori serapan umum. Data tersebut berkaitan dengan penggunaan yang tidak sesuai dengan ejaan baku EYD Edisi Kelima. Walaupun data yang ditemukan sedikit, tetapi kasus seperti itu menjadi perhatian khusus dalam konteks akademik, yang mana kecakapan berbahasa menjadi hal utama.

Temuan pertama terdapat pada kata “jaman” yang seharusnya ditulis “*zaman*” , yang mana kata tersebut merupakan bentuk tidak baku yang masih digunakan karena kebiasaan lama atau pengaruh dalam pengajaran sehari-hari. Data tersebut ditemukan pada kalimat “Interpretasi dari jaman metodologis telah menyumbangkan pada kolonial...” yang menurut kaidah EYD bentuk tidak baku dari bahasa Arab yang sudah dibakukan; zaman, meskipun telah ditetapkan.

Temuan kedua yaitu penggunaan kata “kontuksi” (muncul satu kali) dalam kalimat “...maka tipe kontruksi frasa endosentrik...”. Seharusnya, kata tersebut ditulis konstruksi, sesuai dengan bentuk baku dalam EYD. Kesalahan tersebut bisa diindikasikan sebagai kesalahan dalam mengetik atau fonetik umum; penukaran posisi huruf “n” dan “t”. Namun, kesalahan tersebut tidak boleh diabaikan karena terjadi dalam kajian linguistik, yang mana presisi terhadap istilah sangat krusial.

Kedua temuan tersebut menunjukkan bahwa beberapa penulis masih belum konsisten dalam menggunakan bentuk serapan yang telah dibakukan. Ketidaktepatan yang terjadi dapat memengaruhi kejelasan tulisan dan menurunkan kredibilitas dalam ranah akademik. Oleh karena itu, penting bagi penulis artikel ilmiah untuk memahami dan menerapkan kaidah ejaan yang benar, terutama dalam penggunaan istilah serapan yang umum digunakan.

5. Penutup

Berdasarkan hasil analisis korpus terhadap artikel-artikel jurnal ilmiah terindeks SINTA 5 dan 6, dapat disimpulkan bahwa kesalahan ejaan masih menjadi persoalan signifikan yang memengaruhi kualitas kebahasaan dalam publikasi ilmiah di Indonesia.

Penelitian ini menunjukkan bahwa penulis belum sepenuhnya memahami atau menerapkan kaidah Ejaan Yang Disempurnakan (EYD) edisi kelima, khususnya dalam aspek penulisan kata depan, kata turunan, dan unsur serapan. Kondisi ini membuktikan bahwa lonjakan produktivitas riset belum sejalan dengan ketelitian linguistik dalam penyusunan naskah ilmiah.

Dengan demikian, dapat ditegaskan bahwa peningkatan kualitas publikasi tidak cukup hanya melalui kuantitas, melainkan harus dibarengi dengan penguatan kompetensi kebahasaan. Pendekatan korpus digital juga bisa sebagai alternatif yang terbukti efektif untuk mengidentifikasi pola kesalahan secara sistematis dan dapat menjadi rujukan bagi editor, penulis, dan lembaga pengelola jurnal dalam menyusun strategi peningkatan mutu kebahasaan publikasi ilmiah.

Daftar Pustaka

- Abdugafurovna, R. A., Mukaddas Tursunaliyeva, A., & Turdiyeva, X. Z. (2020). Representation Of Linguistic Issues In Corpus. *Journal of Critical Reviews* . <https://doi.org/10.31838/jcr.07.12.20>
- Awagu, I. V. (2021). Language in Academic Writing: Features and Topical Issues. *Linguistics and Literature Studies*, 9(2), 49–56. <https://doi.org/10.13189/lis.2021.090201>
- Bednarek, M., & Carr, G. (2021). Computer-assisted digital text analysis for journalism and communications research: introducing corpus linguistic techniques that do not require programming. *Media International Australia*, 181(1), 131–151. <https://doi.org/10.1177/1329878X20947124>
- Creswell, J. W. , author. (2017). *Research design : Pendekatan metode kualitatif, kuantitatif, dan campuran*. <https://lib.ui.ac.id>
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and Conducting Mixed Methods Research* (3 ed.). SAGE Publication, Inc.
- Fry, C. V., Lynham, J., & Tran, S. (2023). Ranking researchers: Evidence from Indonesia. *Research Policy*, 52(5), 104753. <https://doi.org/10.1016/j.respol.2023.104753>
- Herwanis, D., Alianur, M., Maulida, & Nasir, K. (2025). Quantitative Research: Data Collection and Data Analysis for Beginner Researcher in Discussion. *Educompassion: Jurnal Integrasi Pendidikan Islam dan Global*, 2(1), 1–18. <https://doi.org/10.63142/educompassion.v1i3.88>

- Iryani, E. (2025). *Dataset A Digital Corpus-Based Study Of Spelling Errors In Sinta Indexed Articles*. figshare. <https://doi.org/https://doi.org/10.6084/m9.figshare.29183660.v1>
- Kemenristek-Dikti. (2018). *Peraturan Menteri Riset, Teknologi, Dan Pendidikan Tinggi Republik Indonesia Nomor 9 Tentang Akreditasi Jurnal Ilmiah*. www.peraturan.go.id
- Khan, A., & Rasul, S. (2020). Extraction of Semantic Domains Through Corpus Tools. *Global Language Review*, *V*(1), 153–168. [https://doi.org/10.31703/glr.2020\(V-1\).17](https://doi.org/10.31703/glr.2020(V-1).17)
- Liang, F. (2023). Book Review on *Corpus Linguistics and Translation Tools for Digital Humanities: Research Methods and Applications*, by Stefania M. Maci, Michele Sala, Bloomsbury, 2022, Xiv+249 Pp. *Journal of China Computer-Assisted Language Learning*, *3*(2), 362–368. <https://doi.org/10.1515/jccall-2023-0007>
- Marselina, S. (2022). Analisis Kesalahan Ejaan Bahasa Indonesia pada Artikel Ilmiah Mahasiswa STIE Alam Kerinci. *Sintaks: Jurnal Bahasa & Sastra Indonesia*, *2*(1), 101–106. <https://doi.org/10.57251/sin.v2i1.272>
- Mashabi, S., & Kasih, A. P. (2024). Kualitas Publikasi Ilmiah Indonesia Dikritik, Kemendikbud: Pengaruhi Reputasi Dosen. *Kompas.com*. <https://www.kompas.com/edu/read/2024/03/01/075313871/kualitas-publikasi-ilmiah-indonesia-dikritik-kemendikbud-pengaruhi-reputasi>
- Meyer, C. F. (2023). *English Corpus Linguistics*. Cambridge University Press. <https://doi.org/10.1017/9781107298026>
- Nainggolan, I. C., Fasyah, N., Pangabea, N. P., Pardosi, N. M. V., Saragih, Y. V., & Hadi, W. (2024). Analisis Kesalahan Penggunaan Bahasa Sesuai dengan Ejaan yang disempurnakan Edisi V pada Penulisan Karya Ilmiah Mahasiswa Universitas Negeri Medan. *Jurnal Bahasa Daerah Indonesia*, *1*(2), 9–9. <https://doi.org/10.47134/JBDI.V1I2.2502>
- Nurizka, A., Putri P, N., Himawan P, R., & Ulya, C. (2021). Telaah Kesalahan Berbahasa Indonesia Pada Jurnal Bahasa Dan Sastra Indonesia Universitas Negeri Semarang. *Jurnal Edukasi Khatulistiwa : Pembelajaran Bahasa dan Sastra Indonesia*, *4*(2), 89–98. <https://doi.org/10.26418/EKHA.V4I2.44295>
- Nurizka R, A., Putri P, N., Prasetyo, R. H., & Ulya, C. (2021). Telaah Kesalahan Berbahasa Indonesia Pada Jurnal Bahasa Dan Sastra Indonesia Universitas Negeri Semarang. *Jurnal Edukasi Khatulistiwa : Pembelajaran Bahasa dan Sastra Indonesia*, *4*(2), 89. <https://doi.org/10.26418/ekha.v4i2.44295>
- Politzer-Ahles, S., Girolamo, T., & Ghali, S. (2020). Preliminary evidence of linguistic bias in academic reviewing. *Journal of English for Academic Purposes*, *47*, 100895. <https://doi.org/10.1016/j.jeap.2020.100895>

- Qoiriah, A. (2021). Implementasi Algoritma Jaro-Winkler Distance dan N-Gram untuk Deteksi dan Prediksi Perbaikan Kesalahan Penulisan Kata Bahasa Indonesia pada Karya Tulis Ilmiah Mahasiswa. *Journal of Informatics and Computer Science (JINACS)*.
https://www.academia.edu/83042194/Implementasi_Algoritma_Jaro_Winkler_Distance_dan_N_Gram_untuk_Deteksi_dan_Prediksi_Perbaikan_Kesalahan_Penulisan_Kata_Bahasa_Indonesia_pada_Karya_Tulis_Ilmiah_Mahasiswa
- Rahmanian, S., & Kuperman, V. (2019). Spelling Errors Impede Recognition of Correctly Spelled Word Forms. *Scientific Studies of Reading*, 23(1), 24–36.
<https://doi.org/10.1080/10888438.2017.1359274>
- Rayson, P., Mariani, J., Anderson-Cooper, B., Baron, A., Gullick, D., Moore, A., & Wattam, S. (2016). Towards Interactive Multidimensional Visualisations for Corpus Linguistics. *Journal for Language Technology and Computational Linguistics*, 31(1), 27–49. <https://doi.org/10.21248/jlcl.31.2016.200>
- Saddler, B. (2023). *It Was a Dark and Stormy Sentence: Teaching the Fine Art of Sentence Construction* (hlm. 147–166). https://doi.org/10.1007/978-3-031-36472-3_9
- Serungke, M., Lutfiah, A., Fadillah, M. A., Rambe, N. B., & Maulani, S. (2023). Analisis Kesalahan Penerapan Ejaan Bahasa Indonesia Pada Jurnal Literasi: Jurnal Ilmiah Pendidikan Bahasa, Sastra Indonesia dan Daerah. *EUNOIA (Jurnal Pendidikan Bahasa Indonesia)*, 3(1), 10.
<https://doi.org/10.30821/EUNOIA.V3I1.2466>
- Syahputra, H. (2006). *Perangkat Lunak Identifikasi dan Koreksi Kesalahan Berbahasa Indonesia Pada Karya Ilmiah*.
https://www.academia.edu/37982252/Perangkat_Lunak_Identifikasi_dan_Koreksi_Kesalahan_Berbahasa_Indonesia_Pada_Karya_Ilmiah
- Wachidah, M. (2019). Analisis Kesalahan Berbahasa Dalam Brosur Ppdb (Penerimaan Peserta Didik Baru) SMA Sederajat Kabupaten Boyolali. *Jurnal Edukasi Khatulistiwa: Pembelajaran Bahasa dan Sastra Indonesia*, 2(2), 1.
<https://doi.org/10.26418/ekha.v2i2.33025>
- Wan Jusoh, W. N. A. S., Jobar, N. A., Md Yusoff, M. Z. N., & Mahat, H. (2024). Exploring corpus linguistics via computational tool analysis: key finding review. *Indonesian Journal of Electrical Engineering and Computer Science*, 34(2), 1052.
<https://doi.org/10.11591/ijeecs.v34.i2.pp1052-1062>