HUMANIS

p-ISSN: 2528-5076, e-ISSN: 2302-920X
Terakreditasi Sinta-3, SK No: 105/E/KPT/2022
Vol 30.1. Pebruari 2026: 62-80

Luaran Penelitian Non-Tradisional dalam Linguistik di Era Digital: Studi Kasus *VerbInd* Pangkalan Data Verba Bahasa Indonesia Berbasis Korpus

Non-Traditional Linguistic Research Output in the Digital Era: A Case Study of VerbInd *a Corpus-Based Database of Indonesian Verb*

Gede Primahadi Wijaya Rajeg
Program Studi Sastra Inggris, *Computer-assisted Lexicology and Lexicography* (CompLexico) research group, & *Centre for Interdisciplinary Research on the Humanities and Social Sciences* (CIRHSS), Fakultas Ilmu Budaya, Universitas Udayana
Email korespondensi: primahadi_wijaya@unud.ac.id

Info Artikel

Masuk: 10 Oktober 2025
Revisi: 13 Desember 2025
Diterima: 23 Desember 2025
Terbit: 28 Pebruari 2026

Keywords: *digital humanities; computational linguistics; corpus linguistics; quantitative linguistics; linguistic database*

Kata kunci: *humaniora digital; linguistik komputasional; linguistik korpus; linguistik kuantitatif; pangkalan data linguistik*

Corresponding Author:
Gede Primahadi Wijaya Rajeg, primahadi_wijaya@unud.ac.id

DOI:
https://doi.org/10.24843/JH.2026.v30.i01.p05

Abstract

The 21st-century research ecosystem highlights the importance of non-traditional outputs such as datasets and software, alongside articles and books. In the Humanities, digital technologies enable data to be presented more openly through online databases. This paper introduces VerbInd*, an open-access online database of morphologically complex verbs in Indonesian, built from corpus texts and processed through computational morphological parsing with further manual verification and editing. The paper demonstrates the use of* VerbInd *in quantitative research on Indonesian verbal morphology and morphosyntax (particularly voice alternation), resulting in traditional outputs, namely book chapters. It also briefly compares* VerbInd *with the sixth edition of the online* Kamus Besar Bahasa Indonesia *(*KBBI*), emphasizing their potential complementarity as linguistic research resources.*

Abstrak

Ekosistem penelitian abad ke-21 menekankan pentingnya luaran non-tradisional seperti dataset dan perangkat lunak, selain artikel dan buku. Dalam bidang Humaniora, teknologi digital mendukung penyajian data secara terbuka melalui pangkalan data daring. Artikel ini memperkenalkan VerbInd*, pangkalan data daring terbuka untuk verba berimbuhan bahasa Indonesia, yang dibangun dari teks korpus dengan penguraian struktur-internal morfologis secara komputasional dan penyuntingan manual lanjutan. Pemanfaatannya ditunjukkan dalam penelitian kuantitatif morfologi verba dan morfosintaksis (utamanya alternasi diatesis) bahasa Indonesia, yang menghasilkan luaran tradisional berupa bab buku. Artikel ini juga membandingkan* VerbInd *dengan* Kamus Besar Bahasa Indonesia (KBBI) *daring edisi keenam, serta menekankan potensi keduanya untuk saling melengkapi sebagai sumber daya penelitian linguistik.*

1. PENDAHULUAN

Secara tradisional, mekanisme penilaian hasil penelitian utamanya (i) terbatas pada jenis luaran berupa tulisan jurnal dan buku (dengan tinjauan sejawat) (Engels & Kulczycki, 2022, p. 2; Hurrell, 2023, p. 3), dan (ii) menitikberatkan pada indeksasi bibliometrik (seperti pada *Web of Science* atau *Scopus*, yang tampak berguna untuk ilmu pengetahuan alam [seperti Kimia dan Farmasi/*Medicine*], namun sulit untuk digunakan dalam menilai dampak sosial dari luaran penelitian bidang ilmu Sosial Humaniora[1] (Engels & Kulczycki, 2022, p. 2; Loprieno et al., 2016, pp. 15–16)). Perkembangan termutakhir ekosistem riset abad 21 mulai memperluas (i) cakupan jenis luaran penelitian (yang tidak hanya berupa tulisan ilmiah dalam jurnal dan buku) dan (ii) butir-butir yang perlu dipertimbangkan dalam penilaian performa penelitian (yang tidak semata-mata menomorsatukan indeksasi numerik)[2] (Hurrell, 2023, pp. 4–5; Raff, 2013).

Selain luaran penelitian tradisional seperti artikel dan buku, keberadaan luaran penelitian yang bersifat non-tradisional mulai meningkat dan beragam. Luaran penelitian non-tradisional tersebut secara luas meliputi di antaranya koleksi digital ("*digital collections*"), luaran "*Geographic Information System*" (GIS), bahan audiovisual, pangkalan/set data ("*datasets*"), naskah pracetak ("*preprints*"), kode pemrograman ("*code*"), peranti lunak ("*software*"), atau algoritma ("*algorithm*") (Hurrell, 2023, p. 3). Hasil survei terhadap anggota *Global Research Council* (GRC) menunjukkan bahwa 76% dari responden GRC sudah melaksanakan penilaian terhadap luaran non-tradisional (Curry et al., 2022, p. 35). Jenis luaran non-tradisional yang paling sering disebutkan dalam survei tersebut adalah kode pemrograman/peranti lunak/algoritma (65%), yang diikuti oleh pangkalan data di peringkat kedua (58%) (Curry et al., 2022, pp. 5, Figure 4). Kedua jenis luaran non-tradisional paling dominan tersebut juga mulai berkembang di bidang Humaniora, khususnya Linguistik, yang akan dijelaskan selanjutnya.

Makalah ini menampilkan satu contoh luaran penelitian non-tradisional dalam linguistik, yaitu *VerbInd* (Pangkalan Data Verba Bahasa Indonesia Berbasis Korpus) (Rajeg & Denistia, 2023b). Penjelasan yang dicakup (Bagian 3.2) adalah fitur-fitur interaktif daring dan informasi yang terkandung dalam *VerbInd* (yaitu frekuensi bentuk verba menyeluruh dan berdasarkan tahun dan ragam teks/*genre* dan struktur morfologis dan sintaksis internal verba, seperti afiks dan kelas kata). Selanjutnya, Bagian 3.3 makalah ini juga secara ringkas mengulas perbedaan antara *VerbInd* dan *Kamus Besar Bahasa Indonesia* (KBBI) daring versi ke-6 (perhatikan Gambar 7). Tahapan dalam menyusun *VerbInd* juga akan diulas terlebih dahulu, beserta ulasan jenis luaran penelitian non-tradisional Linguistik lainnya yang berbeda dengan *VerbInd* (lihat Bagian 2).

Bagian 3.1 akan menjelaskan bahwa *VerbInd* awalnya dirancang khusus untuk menjawab pertanyaan penelitian dalam bidang morfologi, yaitu produktivitas afiksasi verba bahasa Indonesia (perhatikan misalnya Gambar 2). Hasil penelitian tersebut kemudian dituliskan menjadi luaran tradisional (artikel bab buku bunga rampai) namun menjadi entitas terpisah dari pangkalan data yang mendampinginya (Rajeg & Denistia, 2023a). Selanjutnya, Bagian 3.4 akan menunjukkan bahwa data kuantitatif dalam *VerbInd* dapat diperluas guna menjawab pertanyaan dalam bidang morfosintaksis

---

[1] https://enressh.eu/working-group-3/objectives/

[2] Untuk informasi lebih lanjut, kunjungi juga laman *San Francisco Declaration on Research Assessment* (DORA) https://sfdora.org/read/, *European Network for Research Evaluation in the Social Sciences and Humanities* (ENRESSH), dan *Leiden Manifesto* (Hicks et al., 2015) (unduh naskah utuh *Leiden Manifesto* di sini).

bahasa Indonesia (Rajeg & Artawa, 2024), utamanya interaksi antara konstruksi gramatikal (yaitu konstruksi diatesis morfologis bahasa Indonesia) dan unsur leksikal (yaitu verba yang muncul dalam konstruksi diatesis morfologis tersebut). Berkaitan dengan perluasan penggunaan *VerbInd* pada ranah morfosintaksis, *VerbInd* juga dapat digunakan untuk mengkaji lebih lanjut profil morfologis suatu akar kata yang bisa dikaitkan dengan konsep teoretis tertentu. Misalnya, informasi kuantitatif dalam *VerbInd* dapat digunakan untuk mengukur ketiadaan signifikan (secara statistik) suatu bentuk morfologi dari suatu akar verba, yang dapat mencerminkan konsep **bukti negatif** (*negative evidence*) suatu struktur lingual (Bowerman, 1988; Stefanowitsch, 2006) (perhatikan Tabel 5). Implikasi pedagogis *VerbInd* dalam pembelajaran bahasa Indonesia untuk penutur asing juga diulas secara ringkas di akhir Bagian 3.4.

2. METODE DAN TEORI

Perkembangan Humaniora Digital (*Digital Humanities*) (Drucker, 2021; Warwick et al., 2012) menjadi penyokong pesatnya keberadaan pangkalan data digital dalam bidang Humaniora (EAMENA database, 2023; Kirby et al., 2016; Plutniak, 2025; Watts et al., 2015). *VerbInd* (https://gederajeg.github.io/database-verba-bahasa-indonesia/) sebagai suatu luaran penelitian non-tradisional dalam bidang Linguistik mengejawantahkan tiga pilar utama dalam Humaniora Digital, yaitu (i) bahan (*materials*), (ii) pengolahan (*processing*), dan (iii) penyajian (*presentation*) (Drucker, 2021, pp. 3–5; Rajeg, 2025a). Bagian ini akan menjelaskan dua dari tiga pilar utama tersebut dan menjelaskan pilar "penyajian" pada Bagian 3. Secara ringkas, aspek pertama, yaitu "bahan", dalam hal ini merujuk pada sumber teks; aspek kedua ("pengolahan") melibatkan unsur komputasional yang tercermin pada repositori terbuka pengolahan bahan untuk *VerbInd* (Rajeg & Denistia, 2023c). Terakhir, aspek "penyajian" memanfaatkan wahana daring (*online platform*) yang memungkinan interaktifitas (Drucker, 2021, pp. 3–5).

Data yang terkandung dalam *VerbInd* berasal dari *Korpus Tata Bahasa Indonesia Kontemporer* (Korpus TBIK) sebagai sumber teksnya. Korpus ini dirancang oleh *Pusat Pengembangan dan Pelindungan Bahasa dan Sastra, Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi*. Perancangan Korpus TBIK merupakan bagian dari penyusunan tata bahasa Indonesia berbasis korpus yang diprakarsai oleh instansi yang sama[3]. Untuk saat ini, Korpus TBIK belum diluncurkan secara terbuka untuk khalayak umum, namun tersedia untuk penulis kontributor tata bahasa Indonesia kontemporer.

Korpus TBIK secara total berukuran 18.103.443 kali kemunculan kata. Komposisi teks yang terkandung dalam korpus TBIK terbagi menjadi 12 kategori teks. Tabel 1 menampilkan ukuran tiap-tiap kategori teks.

[3] Tautan rekaman seminar hasil untuk topik morfologi https://www.youtube.com/live/fD70Nhs6vug?si=XTR1Zb3rGyfTKiCl.

Tabel 1: Ukuran Korpus TBIK berdasarkan kategori teks (*genre*)

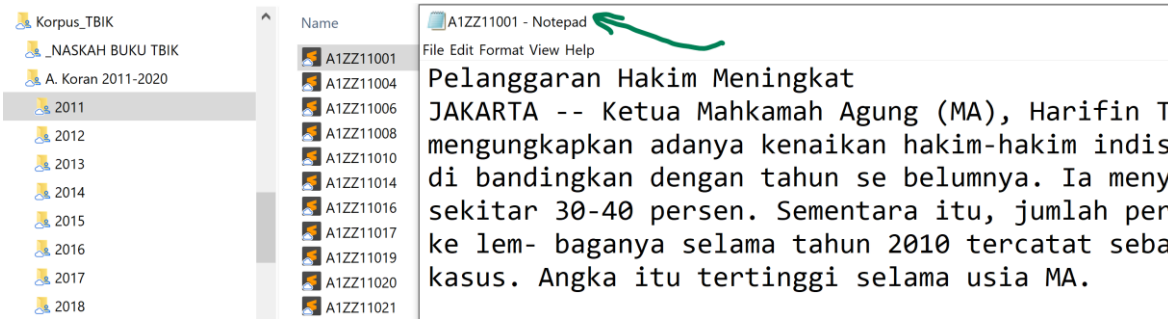
KATEGORI TEKS	UKURAN
Koran	1.017.422
Majalah	1.697.870
Cerpen	1.175.226
Novel	1.253.595
Buku_Teks	1.707.569
Jurnal	1.637.712
Disertasi_Tesis_Skripsi	1.595.030
Biografi	1.644.005
Populer	1.740.642
Perundang_undangan	1.530.914
Laman_Resmi	1.354.123
Surat_Resmi	1.749.335

Selain informasi kategori teks, Korpus TBIK juga mengandung informasi tahun yang terentang dari tahun 2011 hingga 2020 (yaitu periode 10 tahun; perhatikan Tabel 2).

Tabel 2: Ukuran Korpus TBIK berdasarkan tahun/periode

TAHUN	UKURAN
2011	1.642.502
2012	1.668.959
2013	1.760.784
2014	1.649.492
2015	1.769.631
2016	1.720.577
2017	1.957.005
2018	2.016.853
2019	2.026.664
2020	1.890.976

Secara total, terdapat 13.305 berkas teks (dalam format *plain text* [.txt] (Gambar 1)) dari keseluruhan ragam teks dan tahun yang membentuk korpus TBIK.



Gambar 1: Nukilan salah satu berkas teks (`A1ZZ11001.txt`) (sub-)korpus TBIK ragam Koran yang dibuka melalui peranti *Notepad* di Windows.

Dengan besarnya jumlah kata dan berkas korpus TBIK tersebut, pemrosesan secara progamatik perlu dilakukan. Penulis merancang kode pemrograman (Rajeg & Denistia, 2023c) untuk memproses berkas korpus TBIK tersebut menggunakan bahasa pemrograman R (R Core Team, 2025); kode tersebut tersedia terbuka pada https://github.com/gederajeg/afiksasi-verba-bahasa-indonesia/blob/main/1-pre-processing-corpus.R. Data korpus TBIK yang telah diolah guna membangun *VerbInd* dan digunakan dalam makalah ini tersedia terbuka pada https://doi.org/10.5281/zenodo.7812619 (Rajeg & Denistia, 2023c).

Tahap pertama pengolahan data pra-analisis adalah menghasilkan daftar frekuensi kata (dari bentuk teks paragraf seperti pada Gambar 1) untuk tiap ragam teks dan tiap tahun/periode. Tahap utama kedua adalah melakukan analisis (struktur internal) morfologis tiap-tiap kata, utamanya verba yang menjadi fokus penelitian. Pemecahan struktur internal kata secara morfologis awalnya dilakukan secara komputasional programatik menggunakan *MorphInd: Indonesian Morphological Analyzer* (Larasati et al., 2011)[4]. Tabel 3 menampilkan nukilan luaran analisis *MorphInd* untuk kata berawalan *meN-*.

Tabel 3: Sampel luaran analisis MorphInd

KATA	MORPHIND
membutuhkannya	meN+butuh<v>+kan_VSA+dia<p>_PS3
menghilangkannya	meN+hilang<a>+kan_VSA+dia<p>_PS3
mengajar	meN+ajar<v>_VSA
menerawang	meN+terawang<n>_VSA
menghasut	meN+hasut<v>_VSA

Label “VSA” pada Tabel 3 adalah singkatan dari *verb* (V), *singular* (S), dan (diatesis) *active* (A). Sedangkan label “PS3” merujuk pada *personal pronoun* (P), *singular* (S), dan *third person* (3). Huruf kecil dalam “<.>” merujuk pada label kelas kata akar kata[5]. *MorphInd* tidak selalu berhasil mengenali dan memecah struktur internal kata seperti ditunjukkan pada Tabel 4; luaran ini ditandai dengan label “_X–” dan “<x>”.

Tabel 4: Luaran MorphInd yang tidak mengenali unsur pembentuk katanya

KATA	MORPHIND
mengelem	mengelem<x>_X–
menjiplak	menjiplak<x>_X–
berkeluh-kesah	ber+keluh<n>_VSADASHkesah<x>_X–
ber-bicara	ber<x>_X–DASHbicara<v>_VSA
mengontaminasi	mengontaminasi<x>_X–

Untuk kasus seperti pada Tabel 4, penyuntingan dan identifikasi struktur internal kata dilakukan secara manual terhadap luaran morphind tersebut[6].

[4] Laman utama *MorphInd* adalah https://septinalarasati.com/morphind/.

[5] Daftar *tagset* lengkap dalam *MorphInd* dapat dilihat melalui tautan https://septinalarasati.com/morphind/#tagset

[6] Berkas kode pemrograman R yang mencatat perubahan dan penyuntingan yang dilakukan tersedia pada repositori GitHub (Rajeg & Denistia, 2023c); perhatikan berkas kode R dengan inisial no `3-1-..`, `4-..`, dan `6-1-...`.

Penting diperhatikan bahwa pada bidang Linguistik, keberadaan luaran penelitian non-tradisional seperti, dan yang menginspirasi, *VerbInd*, mulai meningkat. Beberapa contohnya adalah set data/pangkalan data leksikal (Bond et al., 2014; Kaiping & Klamer, 2018; Krauße et al., 2024; List et al., 2022; Rajeg et al., 2024); pangkalan data morfologi (Denistia, 2023, untuk nomina; Nomoto et al., 2018, untuk kamus morfologi, namun tidak mengandung informasi frekuensi, ragam teks, dan tahun seperti yang disajikan oleh *VerbInd*); ataupun pangkalan data pola gramatikal (Haspelmath, 2025). Peningkatan pangkalan data tersebut juga sejalan dengan perkembangan Sains Terbuka (*Open Science*) (The Royal Society, 2025) dan permasalahan keilmuan secara umum, yaitu keterciptaan ulang sains (yaitu, isu *reproducibility*) (Baker, 2016). Dalam mengikuti perkembangan ekosistem riset terkini, sebuah manifesto terkait cara memberikan penghargaan/kredit (salah satunya dalam hal menyitir) terhadap jenis luaran non-tradisional bidang linguistik tersebut telah dihasilkan (Berez-Kroeker et al., 2018).

Makalah ini ditulis melalui *Quarto* di *RStudio* guna menggabungkan teks biasa dan kode pemrograman R dalam satu wadah. Berkas penulisan artikel tersebut, kode pemrograman R, dan data pendukung makalah ini tersedia terbuka pada tautan berikut https://github.com/complexico/verbind-snbsb2025 (Rajeg, 2025/2025).

3. HASIL DAN PEMBAHASAN

3.1 *VerbInd* dan Luaran Tradisional

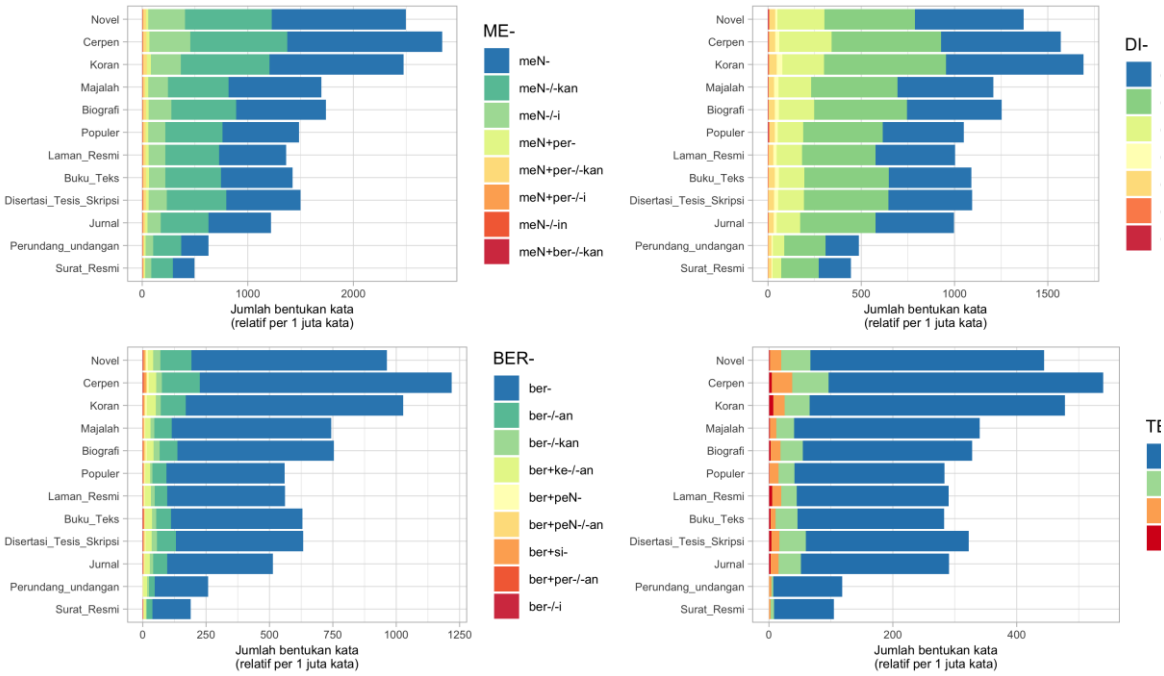
Sebagai luaran non-tradisional berupa pangkalan data daring, *VerbInd* berbeda secara format dari, dan tidak menggantikan namun mendampingi, luaran tradisional tulisan ilmiah bab buku bunga rampai, yang memanfaatkan data di dalam *VerbInd*. Dalam kaitannya dengan luaran penelitian tradisional, penyiapan pangkalan data leksikal bahasa Indonesia dari Korpus TBIK, seperti *VerbInd*, ditujukan untuk menjawab pertanyaan penelitian pada ranah morfologi bahasa Indonesia, salah satunya adalah produktivitas afiksasi verba (Rajeg & Denistia, 2023a) dan nomina (Denistia & Rajeg, 2023)[7]. Kajian produktivitas morfologi (lihat Baayen, 2009) dapat dijawab secara lebih sistematis dan terukur dengan keberadaan pangkalan data raya verba berimbuhan bahasa Indonesia, seperti yang diperoleh dari korpus Tata Bahasa Indonesia Kontemporer (TBIK). Keberadaan korpus TBIK memperkaya kajian produktivitas afiksasi karena mengikutkan dimensi (i) jenis teks (Tabel 1; Gambar 2) (Baayen, 1994; 2009, p. 908; Plag et al., 1999) dan (ii) periode/diakronis (Tabel 2) (Baayen, 2009, p. 909; Palmer, 2015), dua variabel tambahan yang dimiliki oleh korpus TBIK, selain ukurannya yang terbilang besar.

Secara ringkas, Gambar 2 menunjukkan tren serupa di antara empat afiksasi verba BI utama terkait produktivitas sub-tipe afiksasinya di ke-12 ragam teks (lihat Rajeg & Denistia, 2023a, untuk ulasan lebih rinci). Yaitu, produktivitas afiksasi verba terbilang paling rendah pada teks dengan ragam resmi, utamanya dokumen perundang-undangan dan surat resmi; sebaliknya, ragam teks yang memiliki produktivitas afiksasi paling tinggi adalah teks-teks fiksi utamanya novel dan cerita pendek (cerpen), selain juga surat kabar. Produktivitas masing-masing sub-tipe afiksasi verba ditunjukkan oleh warna berbeda dari diagram batangnya.

Penulis berpandangan bahwa perbedaan tingkat produktivitas morfologis yang disajikan mencerminkan karakteristik ragam teksnya untuk tujuan komunikatif berbeda.

[7] Bahan pendukung dan data untuk Denistia & Rajeg (2023) dapat diakses melalui https://osf.io/84zfw/.

Di satu sisi, rendahnya produktivitas afiksasi verbal dalam ranah teks legal (perundang-undangan) dan formal (surat resmi) dapat dipicu oleh sifat konservatif ragam teks ini yang mengandalkan bentuk-bentuk yang telah lazim terpatri (*established*) dan tidak menciptakan istilah/bentuk baru guna menghindari kesalahpahaman. Di sisi lain, tingginya produktivitas morfologis pada ragam fiksi seperti novel dan cerita pendek bisa dipicu oleh pentingnya keragaman dan kreatifitas leksikal ragam teks ini guna melukiskan kompleksitas kejadian dan karakter di dalamnya secara lebih rinci dan kaya.



Gambar 2: Produktivitas sub-tipe afiksasi verba bahasa Indonesia utama berdasarkan jumlah bentukan kata unik (*type frequency*) di 12 ragam teks (Rajeg & Denistia, 2023a, p. 17)

Ketimpangan produktivitas morfologis berdasarkan ragam teks pada dasarnya bukan merupakan sebuah fenomena baru dalam bidang morfologi karena telah ditunjukkan dalam kajian terdahulu dalam bahasa Inggris (Baayen, 1994; 2009, p. 908; Palmer, 2015; Plag et al., 1999) dan Belanda (Baayen & Neijt, 1997). Sebaliknya, keragaman produktivitas morfologis berdasarkan ragam teks adalah kebaruan dalam kajian morfologi bahasa Indonesia, yang tidak diulas dalam buku tata bahasa bahasa Indonesia sebelumnya. Kesenjangan wawasan ini dapat dipahami atas dasar data raya digital (korpus) bahasa Indonesia dengan jenis teks beragam belum tersedia dan masih minimnya pemanfaatan peranti teknologi digital dan pendekatan kuantitatif dalam kajian linguistik Indonesia.

Selanjutnya, luaran dari penelitian morfologi berbasis korpus tersebut tidak semata-mata berupa makalah ilmiah dan wawasan mutakhir terkait morfologi bahasa Indonesia. Keberadaan teknologi dan ruang digital memungkinkan pangkalan data suatu penelitian menjadi publikasi tersendiri guna mendampingi makalah ilmiahnya dan digunakan untuk penelitian lebih lanjut. Bagian 3.2 selanjutnya mengulas fitur-fitur yang saat ini tersedia dalam *VerbInd*.

3.2 Fitur-fitur dalam *VerbInd*

Untuk pangkalan data verba berimbuhan bahasa Indonesia, *VerbInd*, publikasi dilakukan dalam format laman daring terbuka agar dapat diakses masyarakat luas (Rajeg & Denistia, 2023b). Secara teknis, laman *VerbInd* dibangun melalui bahasa pemrograman R (R Core Team, 2025). Bagian ini mengulas fitur yang dimiliki *VerbInd* ketika diakses dalam format laman daring. Bagian 3.4 mengulas akses lanjutan dalam format data mentah *VerbInd* (yaitu format `.txt` sebelum disajikan secara daring), yang dapat diolah lebih lanjut, khususnya untuk menjawab pertanyaan penelitian terkait morfosintaksis bahasa Indonesia.

Versi daring dari *VerbInd* memungkinkan pencarian interaktif kata-kata yang terdapat di pangkalan data tersebut[8]. Sebagai contoh, kita dapat menyaring pangkalan data berdasarkan kriteria tertentu yang dipilih melalui kolom-kolom tertentu. Gambar 3 berikut menunjukkan penyaringan kata-kata berakhiran *-kan* (lihat kolom `suff`, yang merupakan singkatan dari "*suffix*") untuk tahun 2012 (lihat kolom `year`) pada ragam buku teks (kolom `genres`).

Frekuensi verba bahasa Indonesia menurut ragam teks dan tahun

↑ year	form	↓ freq	genres ↑	morph	pref	root	root_POS	suf
2012			Buku_Teks					kan
2012	digunakan	239	Buku_Teks	di+guna<n>+k…	di-	guna	n	kar
2012	menunjukkan	181	Buku_Teks	meN+tunjuk<v…	meN-	tunjuk	v	kar
2012	dilakukan	172	Buku_Teks	di+laku<n>+ka…	di-	laku	n	kar
2012	memberikan	162	Buku_Teks	meN+beri<v>+…	meN-	beri	v	kar
2012	menggunakan	157	Buku_Teks	meN+guna<n>…	meN-	guna	n	kar
2012	meningkatkan	155	Buku_Teks	meN+tingkat<n…	meN-	tingkat	n	kar
2012	menghasilkan	154	Buku_Teks	meN+hasil<n>…	meN-	hasil	n	kar

Gambar 3: Penyaringan data verba berdasarkan kolom *`year`*, *`genres`*, dan *`suff`*

Kemudian, melanjutkan kriteria penyaringan data seperti pada Gambar 3, kita dapat menambah kriteria penyaringan pada kolom lain, misalnya kolom `root_POS` dan `root` (Gambar 4); kolom `root_POS` merujuk pada kategori kelas kata (*Part of Speech*) bentuk akar (*root*) sedangkan kolom `root` adalah bentuk akarnya.

Frekuensi verba bahasa Indonesia menurut ragam teks dan tahun

↑ year	form	↓ freq	genres ↑	morph	pref	root	root_POS	su
2012			Buku_Teks			tahan	a	kan
2012	mempertahankan	33	Buku_Teks	meN+per+taha…	meN-+per-	tahan	a	ka
2012	dipertahankan	11	Buku_Teks	di+per+tahan<…	di-+per-	tahan	a	ka

1–2 of 2 rows Previous 1

Gambar 4: Penyaringan data berdasarkan kolom *`root_POS`* dan *`root`*

Gambar 4 menunjukkan penyaringan data untuk kelas kata **a**jektiva pada kolom `root_POS`, khususnya untuk bentuk *tahan* (kolom `root`). Dari luaran kriteria penyaringan data pada Gambar 3 dan Gambar 4, kita dapat melihat bahwa terdapat dua bentuk (kolom `form`) kata turunan untuk akar kata ajektiva *tahan* yang diikuti akhiran *-kan* beserta frekuensinya (kolom `freq`) pada sub-korpus buku teks di tahun 2012:

[8] Penting untuk diperhatikan bahwa makalah ini tidak menyatakan bahwa *VerbInd* secara utuh telah mencakup semua verba bahasa Indonesia.

bentuk aktif *mempertahankan* secara signifikan lebih kerap digunakan (N=33) dibandingkan bentuk pasif *dipertahankan* (N=11) ($p_{\text{Binomial two-tailed}}$ = 1.26E-03).

Fitur pencarian kedua dari *VerbInd* adalah pencarian data keseluruhan bentuk verba tanpa dipecah berdasarkan tahun dan jenis teks (*genres*). Fitur ini tersedia pada tabel bernama **Frekuensi verba bahasa Indonesia keseluruhan** (Gambar 5) (tabel kedua di bawah tabel berjudul **Frekuensi verba bahasa Indonesia menurut ragam teks dan tahun**).

Frekuensi verba bahasa Indonesia keseluruhan

form	root
	senang
bersenang	senang
bersenang-senanng	senang
disenangi	senang
disenangkan	senang
menyenangi	senang
▸ menyenangkan	senang
menyenang	senang

1–7 of 7 rows Previous 1

Gambar 5: Penyaringan data berdasarkan kolom `root`

Gambar 5 menunjukkan model penjaringan data melalui fitur kedua, utamanya berdasarkan kolom `root` untuk menangkap bentukan verba turunan (muncul pada kolom `form`) dari akar kata *senang*. Kolom `freq` menunjukkan frekuensi kemunculan verba turunan pada kolom `form`.

Selanjutnya, tanda segitiga hitam di sebelah kiri bentuk *menyenangkan* pada Gambar 5 mengindikasikan bahwa frekuensi untuk *menyenangkan* merupakan gabungan dari lebih dari satu bentuk ortografis untuk *menyenangkan*. Tanda segitiga tersebut dapat diklik untuk menampilkan bentuk-bentuk ortografis apa saja yang tercakup (perhatikan area di antara warna hijau stabilo pada Gambar 6).

Frekuensi verba bahasa Indonesia keseluruhan

form	root
	senang
bersenang	senang
bersenang-senanng	senang
disenangi	senang
disenangkan	senang
menyenangi	senang
▾ menyenangkan	senang

orth_var	morph	pref	root_POS	suff
menyenangkan (1137); menyenagkan (1)	meN+senang<a>+kan_VSA	meN-	a	kan

menyenang	senang

Gambar 6: Variasi ortografis suatu bentuk verba

Tampilan setelah mengklik tanda segitiga hitam juga berupa tabel bawahan (*sub-table*) dengan kolom hampir sama dengan tampilan pada Gambar 3 (tanpa kolom `root` yang muncul di tabel utama, dan tanpa kolom `year` dan `genres` yang tidak relevan di tabel ini). Kolom `orth_var` menunjukkan variasi bentuk (untuk verba pada kolom `form` di tabel utama) dan frekuensinya (dalam tanda kurung). Variasi ortografis yang ditampilkan pada Gambar 6 mungkin terjadi akibat kesalahan pengetikan (seperti bentuk *menyenagkan* yang muncul sekali pada Gambar 6). Namun, variasi ortografis juga dapat mencerminkan variasi fonologis yang ditranskripsikan, seperti pada bentuk *berizin* vs. *berijin* (yang pembaca dapat temukan dengan menyaring bentuk akar kata *izin* pada kolom `root`) atau *bergerilya* vs. *bergeriliya* (yang dapat ditemukan dengan menyaring bentuk akar kata *gerilya* pada kolom `root`).

3.3 *VerbInd* dan *Kamus Besar Bahasa Indonesia*

Format *database* seperti *VerbInd* memungkinkan pencarian unsur pembentuk kata yang lebih spesifik, seperti menyaring bentukan kata berdasarkan afiks tertentu (Gambar 3). Sebaliknya, fitur ini belum tersedia pada Kamus Besar Bahasa Indonesia (KBBI) edisi ke-6 versi daring. Pengguna terdaftar untuk KBBI VI daring dapat mencari bentuk afiks tertentu (Gambar 7 (a)) namun ketika afiks tersebut dipilih (Gambar 7 (b)), pengguna akan dibawa ke entri definisi afiks tersebut (Gambar 7 (c)), daripada ditampilkan semua bentukan kata dengan afiks yang dipilih.

(a) Pilih suatu kategori afiks

(b) Bentuk afiks yang dipilih

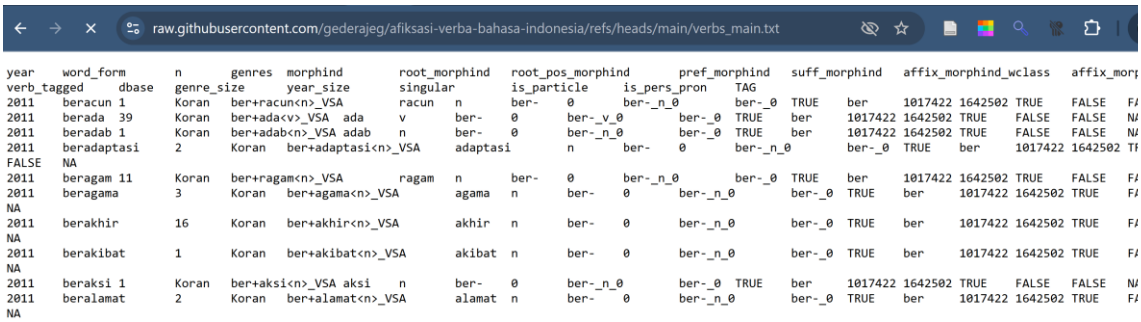
(c) Entri utama suatu afiks

Gambar 7: Pencarian kategori afiks di Kamus Besar Bahasa Indonesia (KBBI) VI Daring

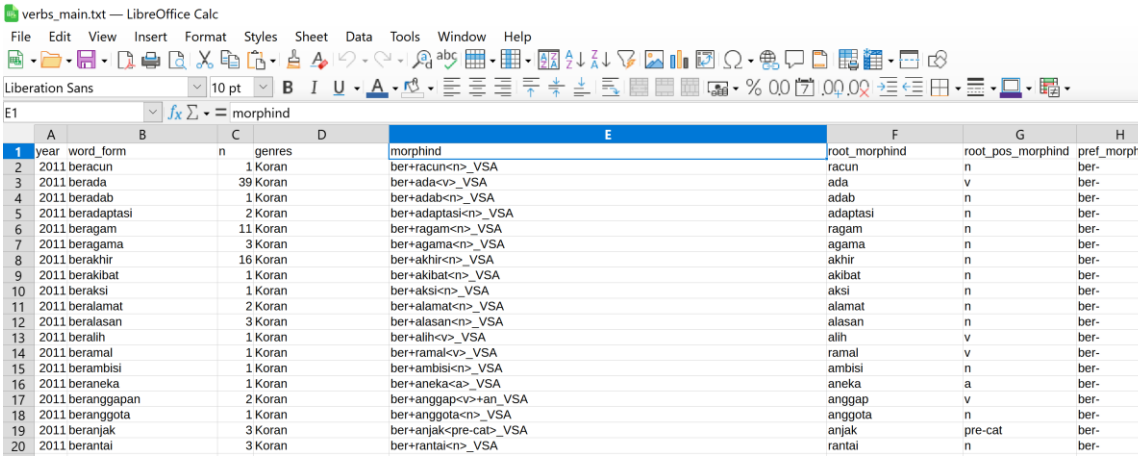
Untuk pengembangan selanjutnya, fitur yang diakomodir model database seperti *VerbInd* (yaitu, mengeluarkan semua bentukan kata dengan afiks tertentu) dapat dipadukan ke dalam fitur kamus daring seperti KBBI VI daring (yang mengandung definisi/fungsi suatu afiks serta contoh-contoh kalimat penggunaan nyata kata turunan dengan afiks tertentu). Pengembangan lainnya adalah mengikutkan informasi frekuensi/kekerapan keseluruhan dan berdasarkan ragam teks untuk suatu bentuk verba berimbuhan ke dalam pangkalan data KBBI VI. Hal ini dapat (i) memperkaya kandungan entri/definisi suatu bentuk verba berdasarkan kelazimannya (kekerapan kemunculannya) secara menyeluruh ataupun berdasarkan ragam tertentu, dan (ii) menambah fungsionalitas (penyortiran hasil) pencarian berdasarkan informasi kekerapan dan ragam teks.

3.4 Penggunaan lebih lanjut (*re-use*) terhadap *VerbInd*

Pembahasan berikutnya mengulas pemanfaatan lebih lanjut dari data yang melandasi *VerbInd*. Pemanfaatan pertama adalah dalam hal penelitian guna menjawab pertanyaan linguistik melalui pendekatan kuantitatif. Data utama *VerbInd* tersedia terbuka dalam format `.txt` atau teks biasa (*plain text*) dengan tiap-tiap kolom dipisahkan oleh penanda `tab` (berkas bernama `verbs_main.txt` pada Rajeg & Denistia, 2023c) (Gambar 8) dan dapat dibuka di program seperti *LibreOffice Calc* (Gambar 8 (b)).



(a) Tampilan ketika dibuka di web



	year	word_form	n	genres	morphind	root_morphind	root_pos_morphind	pref_morph
2	2011	beracun	1	Koran	ber+racun<n>_VSA	racun	n	ber-
3	2011	berada	39	Koran	ber+ada<v>_VSA	ada	v	ber-
4	2011	beradab	1	Koran	ber+adab<n>_VSA	adab	n	ber-
5	2011	beradaptasi	2	Koran	ber+adaptasi<n>_VSA	adaptasi	n	ber-
6	2011	beragam	11	Koran	ber+ragam<n>_VSA	ragam	n	ber-
7	2011	beragama	3	Koran	ber+agama<n>_VSA	agama	n	ber-
8	2011	berakhir	16	Koran	ber+akhir<n>_VSA	akhir	n	ber-
9	2011	berakibat	1	Koran	ber+akibat<n>_VSA	akibat	n	ber-
10	2011	beraksi	1	Koran	ber+aksi<n>_VSA	aksi	n	ber-
11	2011	beralamat	2	Koran	ber+alamat<n>_VSA	alamat	n	ber-
12	2011	beralasan	3	Koran	ber+alasan<n>_VSA	alasan	n	ber-
13	2011	beralih	1	Koran	ber+alih<v>_VSA	alih	v	ber-
14	2011	beramal	1	Koran	ber+ramal<v>_VSA	ramal	v	ber-
15	2011	berambisi	1	Koran	ber+ambisi<n>_VSA	ambisi	n	ber-
16	2011	beraneka	1	Koran	ber+aneka<a>_VSA	aneka	a	ber-
17	2011	beranggapan	2	Koran	ber+anggap<v>+an_VSA	anggap	v	ber-
18	2011	beranggota	1	Koran	ber+anggota<n>_VSA	anggota	n	ber-
19	2011	beranjak	3	Koran	ber+anjak<pre-cat>_VSA	anjak	pre-cat	ber-
20	2011	berantai	3	Koran	ber+rantai<n>_VSA	rantai	n	ber-

(b) Tampilan ketika dibuka di LibreOffice Calc

Gambar 8: Tampilan data utama *VerbInd* (Rajeg & Denistia, 2023c)

Format data seperti *tab-separated plain text* (Gambar 8 (a)) telah diolah secara komputasional untuk analisis linguistik kuantitatif dalam konteks kajian morfologi, utamanya analisis produktivitas morfologis afiksasi verba bahasa Indonesia (Rajeg & Denistia, 2023a) (Gambar 2), yang menjadi tujuan awal *VerbInd* (Bagian 3.1).

Pemanfaatan lain dari data landasan dalam *VerbInd* adalah mengkaji interaksi antara unsur leksikal dan konstruksi gramatikal (Stefanowitsch & Gries, 2009), dalam konteks kajian morfosintaksis kuantitatif bahasa Indonesia. Sebagai contoh, data frekuensi verba dalam *VerbInd* telah diolah lebih lanjut untuk membandingkan preferensi leksikal dua konstruksi diatesis bahasa Indonesia, yaitu diatesis aktif *meN-* dan pasif *di-* (Rajeg & Artawa, 2024)[9]. Untuk tujuan tersebut, Rajeg dan Artawa (2024) menerapkan "Analisis Koleksem Khas" (*Distinctive Collexeme Analysis* [selanjutnya disebut DCA])[10], salah satu ragam teknik kuantitatif di bawah payung metode "Analisis Kolostruksional" (*Collostructional Analysis*) (Gries, 2023; Hilpert, 2014; Stefanowitsch, 2013). DCA dalam hal ini digunakan untuk mengukur ketimpangan statistik penggunaan suatu akar verba dengan diatesis *meN-* dan *di-*. Pengukuran statistik dalam DCA (Rajeg & Artawa, 2024, Sub-bagian 4.3.1) dilakukan untuk semua akar verba yang paling sedikit muncul sekali dengan salah satu dari kedua diatesis, dan pengukuran berulang inilah yang difasilitasi secara komputasional dengan merancang kode pemrograman melalui bahasa pemrograman R.

Hasil dari DCA adalah daftar "koleksem" akar verba yang khas secara statistik untuk suatu bentuk diatesis dibandingkan diatesis lainnya (Rajeg & Artawa, 2024, Tabel 3)[11]. Daftar ini kemudian dapat dianalisis/diinterpretasikan secara kualitatif guna mencirikan perbedaan semantis konstruksi yang dibandingkan, dalam hal ini diatesis aktif dan pasif. Rajeg dan Artawa (2024, Sub-bagian 4.3.2)[12] menemukan bahwa diatesis aktif *meN-* cenderung lebih sering digunakan oleh sekelompok kategori verba

[9] Pemanfaatan *VerbInd* ini disebutkan secara eksplisit pada Bagian 3 dalam Rajeg & Artawa (2024). Kajian ini juga merupakan kerja lanjutan penyusunan Tata Bahasa Indonesia Kontemporer untuk topik "Sintaksis", yang seminarnya dapat disaksikan di kanal *YouTube* melalui tautan ini.

[10] Ulasan *Distinctive Collexeme Analysis* dalam bahasa Indonesia dapat dilihat dalam Rajeg & Artawa (2024, Sub-bagian 4.3) melalui tautan ini, Rajeg (2020), dan Rajeg & Rajeg (2019)

[11] Lihat https://gederajeg.github.io/diatesis-bahasa-indonesia/#tbl-av-vs-di untuk data lebih lengkap terbuka.

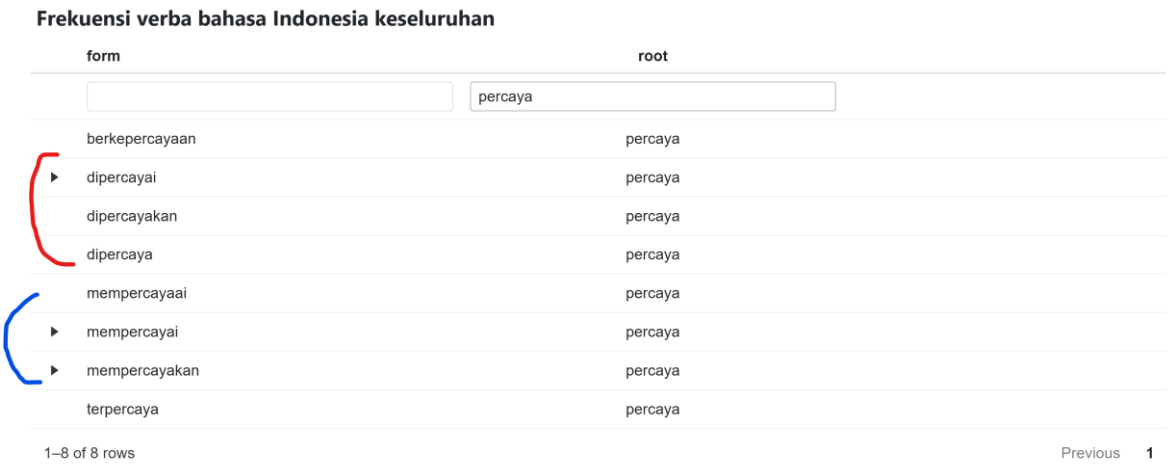
[12] Lihat https://gederajeg.github.io/diatesis-bahasa-indonesia/#sec-dca-av-pass.

yang bersifat 'agentif' dan 'dinamis' (misalnya *tuju*, *cari*, *buat*, *bantu*), sedangkan diatesis pasif *di-* menarik lebih banyak kategori verba yang bersifat 'statis' dan 'mental' (misalnya *isi*, *anggap*, *kenal*, *duga*) (bandingkan temuan kajian serupa dalam bahasa Inggris oleh Gries & Stefanowitsch, 2004).

Data frekuensi *VerbInd* juga dapat digunakan untuk menentukan akar verba yang secara signifikan absen dari suatu konstruksi diatesis. Ketiadaan suatu struktur ini dalam linguistik diistilahkan dengan *negative evidence* (Bowerman, 1988; Stefanowitsch, 2006). Sebagai contoh, perhatikan Tabel 5 terkait data frekuensi verba *percaya* (dan variasinya dengan akhiran *-i* dan *-kan*) pada diatesis aktif *meN-* dan pasif *di-* yang diperoleh dari *VerbInd* (lihat Gambar 9).

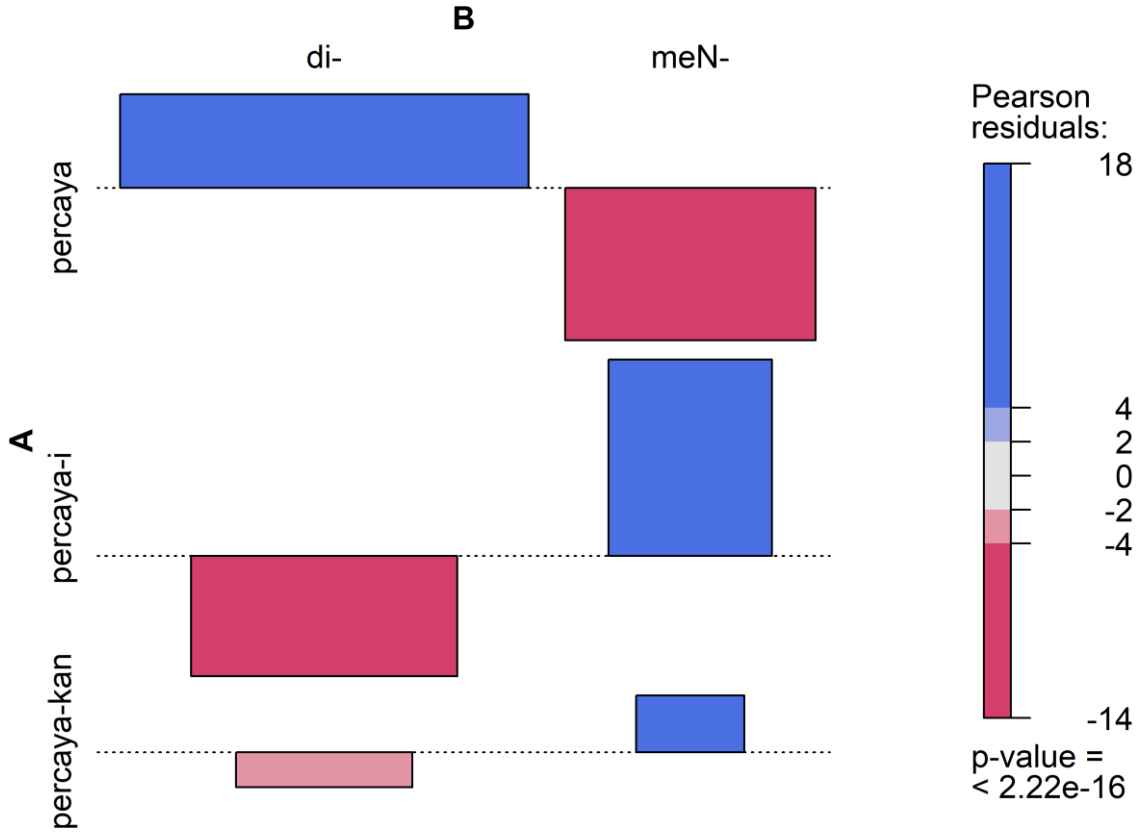
Tabel 5: Frekuensi bentuk diatesis dengan akar verba *percaya* dalam bentuk dasar/tanpa akhiran, diakhiri dengan *-i*, dan *-kan*

	pass. *di-*	av. *meN-*
percaya	746	0
percaya-i	59	258
percaya-kan	68	71



Gambar 9: Nukilan penyaringan data dengan akar kata *percaya* dalam laman *VerbInd*

Dalam pandangan linguistik transformasional (mis. Radford, 1988, p. 420ff), bentuk pasif (bahasa Inggris) merupakan hasil transformasi pergerakan nomina yang dalam struktur batin (*D*[*eep*] *Structure*) adalah objek pos-verbal dari struktur aktif verbanya. Akan tetapi, tidak semua bentuk pasif memiliki realisasi aktifnya, seperti bentuk *dipercaya*, yang frekuensinya paling tinggi pada Tabel 5 dan tidak memiliki pasangan aktifnya, yaitu *mempercaya* (sehingga diberikan nilai frekuensi 0). Seberapa kuat dan signifikan tingkat ketiadaan (atau absennya) konstruksi aktif *mempercaya* ini bisa diukur secara kuantitatif (Stefanowitsch, 2006). Secara statistik, terdapat ketimpangan frekuensi yang signifikan (X^2=785.72; df = 2; p < 0.0001) dan kuat (*Cramer's V* = 0.81) di antara bentuk verba dengan akar kata *percaya* pada Tabel 5. Secara visual, ketimpangan frekuensi signifikan ini ditampilkan pada Diagram Asosiasi pada Gambar 10. Batang biru mengindikasikan asosiasi positif antara bentuk verba dan pemarkah diatesisnya, sedangkan batang kemerahan mengindikasikan disosiasi; semakin gelap warnanya, semakin kuat asosiasi/disosiasinya.



Gambar 10: Diagram asosiasi antara bentuk diatesis pasif *di-* dan aktif *meN-* untuk bentuk dasar *percaya*, *percaya-i*, dan *percaya-kan*

Ketiadaan bentuk *mempercaya* dapat dijelaskan secara tipologi melalui konsep "alternasi diatesis simetrikal" ("*symmetrical voice alternation*"), yang merupakan salah satu ciri unik rumpun bahasa Austronesia (Arka, 2003; Foley, 1998; Himmelmann, 2005; Riesberg, 2014). Dalam alternasi diatesis simetrikal, seperti halnya dalam bahasa Indonesia, bentuk verbanya sama-sama (yaitu secara "simetris") dimarkahi secara morfologis, yaitu dengan awalan *meN-* dan *di-*. Oleh karenanya, secara morfologis, sulit menentukan bentuk tak bermarkah mana yang menjadi bentuk dasar untuk menurunkan bentuk lainnya (Foley, 1998, butir ke-2 bagian 3.3; Himmelmann, 2005, p. 112). Dalam kasus bentuk *di-percaya*, secara morfologis, dan dalam konteks diatesis simetrikal, bentuk tersebut bisa langsung dibangun dengan menggabungkan awalan *di-* dan akar *percaya* tanpa perlu mengasumsikan bahwa bentuk aktif *mem-percaya* (yang tidak ditemukan dalam data sampel) adalah bentuk dasar awalnya. Selanjutnya, secara fungsional dan dari sudut pandang pendekatan Tatabahasa Konstruksional (mis. Croft, 2001; Goldberg, 2013), bentuk *di-percaya* memiliki kemandirian sebagai suatu unit konstruksi dalam menyampaikan fungsi awalan *di-* yang menunjukkan bahwa subjek gramatikal klausanya adalah non-Aktor (yaitu, *Undergoer*) (Arka & Manning, 2008, pp. 62–64) (tanpa perlu dikaitkan dengan bentuk aktif hipotetikal yang tidak ditemukan secara empiris).

Terakhir, selain untuk penelitian, potensi pemanfaatan kedua dari *VerbInd* adalah dalam bidang pengajaran bahasa Indonesia, utamanya untuk penutur asing (BIPA). Sebagai contoh, informasi frekuensi dan bentuk-bentuk yang ditemukan dalam *VerbInd*

dapat digunakan untuk menjelaskan tingkat kelaziman suatu bentuk verba turunan (dan bentuk turunan mana yang berpeluang tidak digunakan) secara lebih terukur objektif berdasarkan data korpus empiris. Contoh lain pemanfaatan temuan analisis produktivitas morfologis verba berasarkan *VerbInd* dalam pengajaran BIPA adalah penentuan afiksasi verba mana yang lebih dulu diajarkan berdasarkan tingkat produktivitasnya. Misalnya, afiksasi *meN+ber-/-kan* dan *meN+per-/-kan* memiliki tingkat produktivitas yang sangat rendah sehingga prioritasnya dalam pengajaran dapat diperkecil dibandingkan bentuk *meN-*, *meN-/-kan*, dan *meN-/-i* yang lebih produktif dan perlu mendapatkan porsi pengajaran dan latihan lebih banyak.

3.5 Pemutakhiran *VerbInd*

VerbInd masih memerlukan sejumlah pemutakhiran ke depannya. Yang pertama, validasi dan penyuntingan manual hasil pemecahan morfologis *MorphInd* masih perlu terus dilanjutkan. Selanjutnya, tampilan web berbasis berkas HTML seperti saat ini perlu diubah menjadi aplikasi daring, misalnya menggunakan wahana *Shiny* melalui bahasa pemrograman R (lih. mis. Krauße et al., 2024). Aplikasi daring ini memungkinkan interaktifitas yang lebih beragam dalam mendukung pengembangan selanjutnya, yaitu menampilkan contoh kalimat dari korpus untuk suatu bentuk verba.

4. SIMPULAN

Ekosistem penelitian abad ke-21 mulai memperluas cakupan jenis luaran penelitian yang dipertimbangkan dalam mekanisme penilaian peneliti(an). Selain luaran tradisional, seperti buku dan artikel ilmiah, luaran non-tradisional, seperti kode pemrograman dan pangkalan data, adalah entitas mandiri yang berbeda dengan, dan dapat mendampingi/dimanfaatkan dalam, luaran tradisional tersebut. Makalah ini mengulas salah satu model dan jenis luaran non-tradisional dalam linguistik, yaitu pangkalan data verba berimbuhan bahasa Indonesia (*VerbInd*) (Bagian 3), dan kaitannya dengan luaran tradisional, yaitu tulisan ilmiah.

Keberadaan dan tahapan penyusunan *VerbInd* (Bagian 2) beriringan dengan tujuan penelitian tradisional penulis, yaitu menjawab rumusan masalah penelitian linguistik (khususnya diawali dengan kajian morfologi (Gambar 2)). Makalah ini juga menunjukkan bagaimana informasi dari pangkalan data yang sama dapat digunakan lebih lanjut (dan diolah secara komputasional kuantitatif) untuk menjawab rumusan masalah penelitian berbeda pada ranah morfosintaksis (Gambar 10)). Kedua penelitian tersebut kemudian dituliskan menjadi luaran tradisional biasa (yaitu artikel bab buku bunga rampai). Makalah ini juga mengulas secara ringkas potensi pemanfaatan *VerbInd* dalam pengajaran (morfologi) verba bahasa Indonesia untuk penutur asing.

Keberadaan teknologi digital-komputasional, perkembangan ekosistem riset terkini, dan meluasnya gerakan Sains Terbuka (*Open Science*) mewadahi penerbitan/diseminasi terbuka dan interaktif sumber data (yang melandasi luaran tradisional) sebagai entitas pangkalan data mandiri. Dari sudut pandang bidang linguistik, perkembangan kajian linguistik mengandalkan data raya dan peranti komputasional memungkinkan luaran penelitian non-tradisional terbuka berorientasi teknologi, seperti pangkalan data, peranti lunak, ataupun tutorial komputasional daring.

UCAPAN TERIMA KASIH

Makalah ini adalah bentuk tulisan dari paparan penulis (Rajeg, 2025b) sebagai (salah satu) narasumber utama pada *Seminar Nasional Bahasa, Sastra, dan Budaya* (SNBSB), 12 September 2025, di Fakultas Ilmu Budaya, Universitas Udayana. Penulis berterima kasih atas masukan dan pertanyaan dari peserta seminar serta Karlina Denistia

(Universitas Sebelas Maret) sebagai kolaborator dalam membangun *VerbInd*. Komentar dan saran membangun dari dua peninjau sejawat, yang penulis sangat hargai, juga telah meningkatkan kualitas dan penyajian makalah ini. Penyelesaian *VerbInd* didukung secara parsial oleh hibah dari *Arts and Humanities Research Council*, UK (AHRC; Nomor Hibah: AH/W007290/1). Rekaman paparan penulis tersedia terbuka pada tautan https://youtu.be/xCaeV2jLvSo.

DAFTAR PUSAKA

Arka, I. W. (2003). Voice systems in the Austronesian Languages of Nusantara: Typology, Symmetricality, and Undergoer Orientation. *Linguistik Indonesia*, *21*(1), 113–139.

Arka, I. W., & Manning, C. D. (2008). Voice and grammatical relations in Indonesian: A new perspective. In P. K. Austin & S. Musgrave (Eds.), *Voice and grammatical relations in Austronesian languages* (pp. 45–69). Center for the Study of Language and Information.

Baayen, R. H. (1994). Derivational productivity and text typology. *Journal of Quantitative Linguistics*, *1*(1), 16–34. https://doi.org/10.1080/09296179408589996

Baayen, R. H. (2009). Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 2, pp. 899–919). Mouton de Gruyter.

Baayen, R. H., & Neijt, A. (1997). Productivity in context: A case study of a Dutch suffix. *Linguistics*, *35*(3), 565–588. https://doi.org/10.1515/ling.1997.35.3.565

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Berez-Kroeker, A. L., Gawne, L., Kung, S. S., Kelly, B. F., Heston, T., Holton, G., Pulsifer, P., Beaver, D. I., Chelliah, S., Dubinsky, S., Meier, R. P., Thieberger, N., Rice, K., & Woodbury, A. C. (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, *56*(1), 1–18. https://doi.org/10.1515/ling-2017-0032

Bond, F., Lim, L. T., Tang, E. K., & Riza, H. (2014). The combined wordnet Bahasa. *NUSA*, *57*, 83–100. https://doi.org/10.15026/79286

Bowerman, M. (1988). The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). B. Blackwell. https://pure.mpg.de/rest/items/item_468143/component/file_532427/content

Croft, W. (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford University Press.

Curry, S., de Rijcke, S., Hatch, A., Pillay, D. (Gansen)., van der Weijden, I., & Wilsdon, J. (2022). *The changing role of funders in responsible research assessment: Progress, obstacles and the way ahead (RoRI Working Paper No.3)* [Report]. Research on Research Institute. https://doi.org/10.6084/m9.figshare.13227914.v2

Denistia, K. (2023). Databases on the Indonesian Prefixes PE- and PEN-. *Journal of Language and Literature*, *23*(1), 13–24. https://doi.org/10.24071/joll.v23i1.4967

Denistia, K., & Rajeg, G. P. W. (2023). Afiksasi nomina dalam bahasa Indonesia. In *Tata bahasa Indonesia kontemporer: Morfologi*. Badan Pengembangan dan Pembinaan Bahasa. https://ora.ox.ac.uk/objects/uuid:7346b8e7-f66b-46ce-8efd-89baac017f6c

Drucker, J. (2021). *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*. Routledge. https://doi.org/10.4324/9781003106531

EAMENA database. (2023). *Sistan: Part 1. Heritage Places* [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.10375902

Engels, T. C. E., & Kulczycki, E. (2022). Introduction: Research assessment in the social sciences. In T. C. E. Engels & E. Kulczycki (Eds.), *Handbook on Research Assessment in the Social Sciences* (pp. 1–6). Edward Elgar Publishing. https://doi.org/10.4337/9781800372559.00006

Foley, W. A. (1998). *Symmetrical Voice Systems and Precategoriality in Philippine Languages* [Paper]. Workshop on Voice and Grammatical Functions in Austronesian Languages, LFG98, Brisbane, Australia. https://doi.org/10.5281/zenodo.5336773

Goldberg, A. E. (2013). Constructionist approaches. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 15–31). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0002

Gries, S. Th. (2023). Collostructional Methods. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (1st ed., pp. 1–6). Wiley. https://doi.org/10.1002/9781405198431.wbeal20003

Gries, S. Th., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, *9*(1), 97–129.

Haspelmath, M. (Ed.). (2025). *CrossGram*. Max Planck Institute for Evolutionary Anthropology. https://crossgram.clld.org/

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, *520*(7548), 429. https://doi.org/10.1038/520429a

Hilpert, M. (2014). Collostructional analysis: Measuring associations between constructions and lexical elements. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 391–404). John Benjamins Publishing Company.

Himmelmann, N. (2005). The Austronesian languages of Asia and Madagascar: Typological characteristics. In K. A. Adelaar & N. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 110–181). Routledge.

Hurrell, C. (2023). Research Assessment Reform, Non-Traditional Research Outputs, and Digital Repositories: An Analysis of the Declaration on Research Assessment (DORA) Signatories in the United Kingdom. *Evidence Based Library and Information Practice*, *18*(4), 2–20. https://doi.org/10.18438/eblip30407

Kaiping, G. A., & Klamer, M. (2018). LexiRumah: An online lexical database of the Lesser Sunda Islands. *PLOS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0205250

Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H.-J., Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., Leehr, D., Low, B. S., McCarter, J., Divale, W., & Gavin, M. C. (2016). D-PLACE: A Global Database of Cultural, Linguistic and Environmental Diversity. *PLOS ONE*, *11*(7), e0158391. https://doi.org/10.1371/journal.pone.0158391

Krauße, D., Rajeg, G. P. W., Pramartha, C. R. A., Zobel, E., Nothofer, B., Hemmings, C., Ogilvie, S., Arka, I. W., & Dalrymple, M. (2024). *EnoLEX: A diachronic*

lexical database for the Enggano language (Version 1.0.0) [Online database]. University of Oxford. https://doi.org/10.25446/oxford.28282169.v1

Larasati, S. D., Kuboň, V., & Zeman, D. (2011). Indonesian morphology tool (MorphInd): Towards an indonesian corpus. *Systems and Frameworks for Computational Morphology*, 119–129. https://doi.org/10.1007/978-3-642-23138-4_8

List, J.-M., Forkel, R., Greenhill, S. J., Rzymski, C., Englisch, J., & Gray, R. D. (2022). Lexibank, a public repository of standardized wordlists with computed phonological and lexical features. *Scientific Data*, *9*(1), 316. https://doi.org/10.1038/s41597-022-01432-0

Loprieno, A., Werlen, R., Hasgall, A., & Bregy, J. (2016). The "Mesurer les Performances de la Recherche" Project of the Rectors' Conference of the Swiss Universities (CRUS) and Its Further Development. In M. Ochsner, S. E. Hug, & H.-D. Daniel (Eds.), *Research Assessment in the Humanities* (pp. 13–21). Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4_2

Nomoto, H., Choi, H., Moeljadi, D., & Bond, F. (2018). MALINDO morph: Morphological dictionary and analyser for Malay/Indonesian. *Proceedings of the LREC 2018 Workshop "the 13th Workshop on Asian Language Resources"*, 36–43. http://lrec-conf.org/workshops/lrec2018/W29/pdf/8_W29.pdf

Palmer, C. C. (2015). Measuring productivity diachronically: Nominal suffixes in English letters, 1400–1600. *English Language and Linguistics*, *19*(1), 107–129. https://doi.org/10.1017/S1360674314000264

Plag, I., Dalton-Puffer, C., & Baayen, H. (1999). Morphological productivity across speech and writing. *English Language and Linguistics*, *3*(2), 209–228. https://doi.org/10.1017/S1360674399000222

Plutniak, S. (2025). *Open-archeOcsean: An interactive catalogue of open source datasets for the archaeology of the Pacific and Southeast Asia regions* (Version 1.0.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.16812839

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radford, A. (1988). *Transformational Grammar: A First Course* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511840425

Raff, J. W. (2013). The San Francisco Declaration on Research Assessment. *Biology Open*, *2*(6), 533–534. https://doi.org/10.1242/bio.20135330

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

Rajeg, G. P. W. (2025a, February). *Languaging in the age of digital humanities: Examples from the Enggano language research*. Fakultas Ilmu Budaya (FIB) Digital Talk (DigiTalk), Fakultas Ilmu Budaya, Universitas Udayana, Bali. https://doi.org/10.25446/oxford.28454975.v1

Rajeg, G. P. W. (2025b, September 12). *Luaran Penelitian Non-Tradisional dalam Linguistik di Era Digital* [Plenary talk]. Seminar Nasional Bahasa, Sastra, dan Budaya (SNBSB). https://doi.org/10.5281/zenodo.17112853

Rajeg, G. P. W. (2025). *Material pendukung dan kode pemrograman R untuk "Luaran Penelitian Non-Tradisional dalam Linguistik di Era Digital"* (Version 1.0.0) [Computer software]. https://doi.org/10.17605/OSF.IO/GN29Q (Original work published 2025)

Rajeg, G. P. W., & Artawa, K. (2024). Kajian korpus kuantitatif terhadap aspek-aspek diatesis dalam bahasa Indonesia. In *Tata bahasa Indonesia kontemporer: Sintaksis*. Badan Pengembangan dan Pembinaan Bahasa. https://doi.org/10.5281/zenodo.10615406

Rajeg, G. P. W., & Denistia, K. (2023a). Afiksasi verba dalam bahasa Indonesia. In *Tata bahasa Indonesia kontemporer: Morfologi*. Badan Pengembangan dan Pembinaan Bahasa. https://ora.ox.ac.uk/objects/uuid:ad496412-7b2d-4b88-b9c1-8b20db52dbac

Rajeg, G. P. W., & Denistia, K. (2023b). *VerbInd: Pangkalan data verba bahasa Indonesia berbasis korpus* (Version 1.0.0) [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7947605

Rajeg, G. P. W., & Denistia, K. (2023c). *Material pendukung untuk Afiksasi Verba Dalam Bahasa Indonesia* (Version 0.0.3) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7812619

Rajeg, G. P. W., Kraußе, D., & Pramartha, C. (2024). EnoLEX: A diachronic lexical database for the Enggano language. In A. Inoue, N. Kawamoto, & M. Sumiyoshi (Eds.), *AsiaLex 2024 proceedings: Asian lexicography - merging cutting-edge and established approaches* (pp. 123–132). https://doi.org/10.25446/oxford.27013864

Rajeg, G. P. W., & Rajeg, I. M. (2019). Analisis Koleksem Khas dan potensinya untuk kajian kemiripan makna konstruksional dalam Bahasa Indonesia. In I. N. Sudipa (Ed.), *ETIKA BAHASA Buku persembahan menapaki usia pensiun: I Ketut Tika* (Vol. 1, pp. 65–83). Swasta Nulus. https://doi.org/10.31227/osf.io/uwzts

Riesberg, S. (2014). *Symmetrical Voice and Linking in Western Austronesian Languages*. De Gruyter Mouton. https://doi.org/10.1515/9781614518716

Stefanowitsch, A. (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory*, *2*(1), 61–77.

Stefanowitsch, A. (2013). Collostructional analysis. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 290–306). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0016

Stefanowitsch, A., & Gries, S. Th. (2009). Corpora and grammar. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 2, pp. 933–951). Mouton de Gruyter.

The Royal Society. (2025). *Open science | Royal Society*. The Royal Society. https://royalsociety.org/journals/open-access/open-science/

Warwick, C., Terras, M., & Nyhan, J. (2012). Introduction. In C. Warwick, J. Nyhan, & M. Terras (Eds.), *Digital Humanities in Practice* (pp. xiii–xx). Facet; published online by Cambridge University Press. https://doi.org/10.29085/9781856049054.001

Watts, J., Sheehan, O., Greenhill, S. J., Gomes-Ng, S., Atkinson, Q. D., Bulbulia, J., & Gray, R. D. (2015). Pulotu: Database of Austronesian Supernatural Beliefs and Practices. *PLoS ONE*, *10*(9), 1–17. https://doi.org/10.1371/journal.pone.0136783